



UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA  
ŠTUDIJ BIOTEHNOLOGIJE

Tjaša TIBAUT

**MODELIRANJE STRUKTURE MOLEKULE TLR15  
IN NAPOVED VEZAVNIH MEST ZA LIGANDE Z  
BIOINFROMACIJSKIMI ORODJI**

MAGISTRSKO DELO

Magistrski študij – 2. stopnja Biotehnologija

Ljubljana, 2014



UNIVERZA V LJUBLJANI  
BIOTEHNIŠKA FAKULTETA  
ŠTUDIJ BIOTEHNOLOGIJE

Tjaša TIBAUT

**MODELIRANJE STRUKTURE MOLEKULE TLR15 IN NAPOVED  
VEZAVNIH MEST ZA LIGАНDE Z BIOINFROMACIJSKIMI  
ORODJI**

MAGISTRSKO DELO  
Magistrski študij – 2. stopnja Biotehnologija

**USE OF BIOINFORMATIC TOOLS FOR TLR15 STRUCTURE  
MODELING AND LIGAND BINDING SITE PREDICTION**

M.SC. THESIS  
Master Study Programmes - Biotechnology

Ljubljana, 2014

Magistrsko delo je zaključek magistrskega študija programa 2. stopnje – Biotehnologija. Opravljeno je bilo v Laboratoriju za imunologijo in celične kulture na Oddelku za zootehniko Biotehniške fakultete Univerze v Ljubljani, v Domžalah.

Študijska komisija študija Biotehnologije je za mentorico diplomskega dela imenovala prof. dr. Mojco Narat in za somentorico dr. Ireno Oven.

Recenzent: prof. dr. Jernej Jakše

Komisija za oceno in zagovor:

Predsednica: prof. dr. Branka JAVORNIK

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo

Članica: prof. dr. Mojca NARAT

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Članica: asist. dr. Irena OVEN

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za zootehniko

Član: prof. dr. Jernej JAKŠE

Univerza v Ljubljani, Biotehniška fakulteta, Oddelek za agronomijo

Datum zagovora:

Magistrsko delo je rezultat lastnega raziskovalnega dela.

Podpisana se strinjam z objavo svoje naloge v polnem tekstu na spletni strani Digitalne knjižnice Biotehniške fakultete. Izjavljam, da je naloga, ki sem jo oddala v elektronski obliki, identična tiskani verziji.

Tjaša TIBAUT

## KLJUČNA DOKUMENTACIJSKA INFORMACIJA

ŠD	Du2
DK	UDK 575.112(043.2)
KG	bioinformatika/homologno modeliranje/podatkovne zbirke/TLR15/ligandi/diacilirani lipopeptidi
AV	TIBAUT Tjaša, dipl. bioteh. (UN)
SA	NARAT Mojca (mentorica), OVEN Irena (somentorica)
KZ	SI-1000 Ljubljana, Jamnikarjeva 101
ZA	Univerza v Ljubljani, Biotehniška fakulteta, Študij biotehnologije
LI	2014
IN	MODELIRANJE STRUKTURE MOLEKULE TLR15 IN NAPOVED VEZAVNIH MEST ZA LIGANDE Z BIOINFORMACIJSKIMI ORODJI
TD	Magistrsko delo (Magistrski študij – 2. stopnja)
OP	XI, 58 str., 11 pregl., 36 sl., 1 pril., 50 vir.
IJ	sl
JI	sl/en
AI	Bioinformatika je hitro se razvijajoča veda, ki s pomočjo računalniških metod in bioinformacijskih orodij zelo olajša rokovanje z ogromno količino podatkov, ki so dandanes na voljo. Ena od metod je homologno modeliranje, ki nam omogoča napovedovanje strukture molekul na podlagi homologije oziroma podobnosti. V magistrski nalogi smo v podatkovnih zbirkah (NCBI in UniProt) poiskali znane sekvence molekule TLR15 in jih med seboj primerjali z orodji za poravnavo sekvenc (BLAST, Clustal Omega). V zbirkah obstaja več kot 60 sekvenc molekule TLR15 različnih vrst ptičev, katerih podobnost je odvisna od sorodnosti organizmov. Za nadaljnje delo smo uporabili sekvenco molekule TLR15 kokoši ( <i>Gallus gallus</i> ). Preden smo se lotili modeliranja smo s pomočjo strežnika Protein Model Portal preverili, če so na spletu že kakšne strukture te molekule. Strežnik ni našel nobenega modela, ki bi ustrezal sekvenci proteina TLR15, zato smo uporabili možnost interaktivnega modeliranja, kjer smo posredno napovedali strukturo s strežniki SWISS-MODEL, HHpred, ModWeb ter Phyre <sup>2</sup> . Slednjega smo uporabili še za neposredno napoved modela molekule. Napovedane modele smo primerjali med sabo, končni model pa analizirali in napovedali potencialne ligande ter njihova vezavna mesta. Iz dobljenih modelov lahko sklepamo, da ima molekula TLR15 značilno solenoidno strukturo z regijami, ki so specifične za ta protein. Glede na rezultate napovedi vezavnih mest lahko sklepamo, da so te specifične regije vezavna mesta za ligand ter da je potencialni ligand molekule diaciliran lipopeptid.

#### KEY WORDS DOCUMENTATION

DN Du2  
DC UDC 575.112(043.2)  
CX bioinformatics/homology modeling/databases/TLR15/ligands/diacylated lipopeptides  
AU TIBAUT, Tjaša  
AA NARAT Mojca (supervisor), OVEN Irena (co-advisor)  
PP SI-1000 Ljubljana, Jamnikarjeva 101  
PB University of Ljubljana, Biotechnical faculty, Master Study in Biotechnology  
PY 2014  
TI USE OF BIOINFORMATIC TOOLS FOR TLR15 STRUCTURE MODELING AND LIGAND BINDING SITE PREDICTION  
DT M. Sc. Thesis (Master Study Programme)  
NO XI, 58 p., 11 tab., 36 fig., 1 ann., 50 ref.  
LA sl  
AL sl/en  
AB Bioinformatics is a fast emerging field in which methods and tools from computer science are used to handle the large amounts of biological data available today. Among these methods is homology modeling, where protein structure is predicted based on structural properties of related molecules. We compared known sequences of the TLR15 molecule found in online databases (NCBI, UniProt) using sequence-alignment tools (BLAST, Clustal Omega). Over 60 TLR15 sequences were found, belonging to various bird species. The similarity of these sequences depends on the evolutionary relationship between their respective species. We focused on the chicken (*Gallus gallus*) TLR15 sequence. No existing structural data for this molecule were found on the Protein Model Portal. We used the SWISS-MODEL, HHpred, ModWeb and Phyre<sup>2</sup> servers to indirectly predict the structure of this protein. Phyre<sup>2</sup> was also used for direct prediction of the molecular model. We compared the predicted structures and analyzed the final model to find potential ligands and their binding sites. Our models show that the TLR15 molecule has a typical solenoid structure with protein-specific regions. Based on the results of binding-site prediction we conclude that these regions potentially serve as binding sites for diacylated lipopeptides.

## KAZALO VSEBINE

	str.
KLJUČNA DOKUMENTACIJSKA INFORMACIJA.....	III
KEY WORDS DOCUMENTATION.....	IV
KAZALO VSEBINE.....	V
KAZALO PREGLEDNIC.....	VIII
KAZALO SLIK.....	IX
KAZALO PRILOG.....	X
SLOVARČEK.....	XI
<b>1 UVOD.....</b>	<b>1</b>
1.1 NAMEN DELA.....	1
<b>2 PREGLED OBJAV.....</b>	<b>2</b>
2.1 KLASIFIKACIJA PROTEINOV.....	2
<b>2.1.1 Proteinske domene.....</b>	<b>2</b>
<b>2.1.2 Ostale klasifikacijske enote.....</b>	<b>2</b>
2.1.2.1 Proteinski motivi.....	2
2.1.2.2 Proteinske ponovitve.....	2
2.1.2.3 Proteinski kompleksi.....	3
<b>2.1.3 Klasifikacija glede na tipe proteinov.....</b>	<b>3</b>
2.2 RECEPTORJI PODOBNI TOLL-U.....	4
<b>2.2.1 Vloga v imunskejem sistemu.....</b>	<b>4</b>
2.2.1.1 Tarče (ligandi) TLR-jev (PAMP-i, ki jih prepoznaajo TLR-ji).....	5
<b>2.2.2 Struktturna biologija TLR-jev.....</b>	<b>5</b>
2.2.2.1 N-terminalna domena.....	6
2.2.2.2 Transmembranska domena.....	6
2.2.2.3 Domena Toll/interlevkin-1 receptor.....	7
<b>2.2.3 Toll-u podobni receptorji pri ptičih.....</b>	<b>7</b>
2.2.3.1 TLR15.....	7
2.3 BIOINFORMATIKA.....	8
<b>2.3.1 Homologno modeliranje.....</b>	<b>8</b>
<b>2.3.2 Metode za sekvenčno-struktурno poravnavo.....</b>	<b>9</b>
2.3.2.1 Sekvenčno-sekvenčne metode.....	10
2.3.2.2 Profilno-sekvenčne in HMM-sekvenčne metode za poravnavo.....	10
2.3.2.3 Profilno-profilne in HMM-HMM metode za poravnavo.....	11
<b>2.3.3 Metode za poravnavo več sekvenc.....</b>	<b>11</b>
<b>2.3.4 Hibridne metode, popolno integrirani avtomatski strežniki in meta-strežniki .</b>	<b>12</b>
2.4 BIOINFORMATIJSKA ORODJA.....	13
<b>2.4.1 Evropski bioinformacijski inštitut.....</b>	<b>13</b>
2.4.1.1 Universal Protein Resource (UniProt).....	13
2.4.1.2 Clustal Omega.....	14
<b>2.4.2 Nacionalni center za biotehnoško informacijo.....</b>	<b>14</b>
2.4.2.1 BLAST.....	14

	str.
<b>2.4.3 Podatkovna baza TollML.....</b>	<b>15</b>
<b>2.4.4 Strežnik Protein Model Portal.....</b>	<b>16</b>
2.4.4.1 HHpred.....	16
2.4.4.2 SWISS-MODEL.....	16
2.4.4.3 ModWeb.....	17
2.4.4.4 Phyre <sup>2</sup> .....	17
<b>2.5 METODE ZA PREPOZNAVANJE VEZAVNIH MEST ZA LIGANDE NA PROTEI- NIH.....</b>	<b>18</b>
<b>2.5.1 Metoda COACH.....</b>	<b>19</b>
<b>3 METODE.....</b>	<b>20</b>
3.1 ANALIZA OBSTOJEČIH ZAPOREDIJ MOLEKULE TLR15 S POMOČJO EVROP- SKEGA BIOINFORMACIJSKEGA INŠTITUTA.....	20
<b>3.1.1 Iskanje ter pridobitev obstoječih zaporedij s podatkovno bazo UniProt.....</b>	<b>20</b>
<b>3.1.2 Poravnava več sekvenc s programom Clustal Omega.....</b>	<b>20</b>
3.2 ANALIZA PODOBNOSTI MED ZAPOREDJI MOLEKULE TLR15 S POMOČJO NACIONALNEGA CENTRA ZA BIOTEHNOLOŠKO INFORMACIJO.....	21
<b>3.2.1 Poravnava sekvenc z algoritmom BLAST.....</b>	<b>21</b>
3.3 ANALIZA MOLEKULE TLR15 Z ORODJEM TOLLML.....	21
<b>3.3.1 Primerjava našega zaporedja z zaporedji v podatkovni zbirki TollML z algorit-         mom WU-BLAST.....</b>	<b>21</b>
3.4 ISKANJE EKSPERIMENTALNIH STRUKTUR IN TEORETIČNIH MODELOV TER INTERAKTIVNO MODELIRANJE S POMOČJO STREŽNIKA PROTEIN MODEL PORTAL.....	22
<b>3.4.1 Iskanje eksperimentalnih struktur in teoretičnih modelov.....</b>	<b>22</b>
<b>3.4.2 Iskanje modelov s podobnimi sekvencami.....</b>	<b>23</b>
<b>3.4.3 Interaktivno modeliranje.....</b>	<b>23</b>
3.5 NEPOSREDNO MODELIRANJE S STREŽNIKOM PHYRE <sup>2</sup> .....	24
3.6 PRIMERJAVA MODELOV ZMODELIRANIH Z RAZLIČNIMI STREŽNIKI TER ANALIZA MODELA MOLEKULE TLR15.....	24
<b>3.6.1 Primerjava modelov zmodeliranih s strežnikom Phyre<sup>2</sup> (posredno s PMP in ne-         posredno).....</b>	<b>25</b>
<b>3.6.2 Primerjava modelov zmodeliranih interaktivno s PMP (HHpred, SWISS-MO-         DEL, ModWeb in Phyre<sup>2</sup>).....</b>	<b>25</b>
<b>3.6.3 Analiza napovedanih domen modela molekule TLR15.....</b>	<b>25</b>
3.7 NAPOVED ŽEPOV, VOTLIN/LUKENJ TER VEZAVNIH MEST ZA LIGANDE.....	26
<b>3.7.1 Napoved vezavnega mesta za ligande z meta-strežnikom COACH.....</b>	<b>26</b>
<b>4 REZULTATI.....</b>	<b>27</b>
4.1 OBSTOJEČA ZAPOREDJA MOLEKULE TLR15 PRIDOBELJENA S POMOČJO PODATKOVNE BAZE UNIPROT.....	27
4.2 PORAVNAVA OBSTOJEČIH SEKVENC MOLEKULE TLR15.....	27
<b>4.2.1 Poravnava več sekvenc s programom Clustal Omega.....</b>	<b>27</b>
<b>4.2.2 Poravnave dveh sekvenc s programom BLAST.....</b>	<b>28</b>
4.3 ANALIZA SEKVENCE MOLEKULE TLR15 S PODATKOVNO BAZO TOLLML	28
<b>4.3.1 Primerjava izbrane sekvence (C4PBP0) z vsemi sekvencami v podatkovni bazi</b>	<b>28</b>

	str.
<b>4.3.2 Domene molekule TLR15 določene s programom LRRFinder.....</b>	<b>29</b>
<b>4.4 EKSPERIMENTALNE STRUKTURE IN TEORETIČNI MODELI MOLEKULE TLR15 IN MOLEKULE S PODOBNIMI SEKVENCAMI.....</b>	<b>31</b>
<b>4.4.1 Eksperimentalne strukture in teoretični modeli na voljo na spletu.....</b>	<b>31</b>
<b>4.4.2 Iskanje modelov s podobnimi sekvencami.....</b>	<b>31</b>
<b>4.5 INTERAKTIVNO (POSREDNO) MODELIRANJE S POMOČJO PMP.....</b>	<b>33</b>
<b>4.5.1 Model izračunan s strežnikom HHpred.....</b>	<b>33</b>
<b>4.5.2 Modeli izračunani s strežnikom SWISS-MODEL.....</b>	<b>33</b>
<b>4.5.3 Model izračunani s strežnikom ModWeb.....</b>	<b>35</b>
<b>4.5.4 Modela izračunana s strežnikom Phyre<sup>2</sup> .....</b>	<b>35</b>
<b>4.6 MODELI IZRAČUNANI NEPOSREDNO S STREŽNIKOM PHYRE<sup>2</sup>.....</b>	<b>37</b>
<b>4.6.1 Model izračunan v načinu »Normal«.....</b>	<b>37</b>
<b>4.6.2 Modeli izračunani v načinu »Intensive«.....</b>	<b>38</b>
<b>4.7 ANALIZA DOMENE LEVCINSKIH PONOVITEV MODELA TLR 15 TER PRI- MERJAVA MODELOV ZMODELIRANIH Z RAZLIČNIMI STREŽNIKI.....</b>	<b>40</b>
<b>4.7.1 Grafični prikaz domene levcinskih ponovitev modela molekule TLR15 (TLR15_6).....</b>	<b>40</b>
<b>4.7.2 Primerjava modelov zmodeliranih s strežnikom Phyre<sup>2</sup>.....</b>	<b>40</b>
<b>4.7.3 Primerjava modelov zmodeliranih interaktivno s PMP (HHpred, SWISS-MO- DEL, ModWeb in Phyre<sup>2</sup>).....</b>	<b>41</b>
<b>4.8 NAPOVED VEZAVNIH MEST ZA LIGANDE.....</b>	<b>42</b>
<b>4.8.1 Vezavna mesta napovedana z metodo COACH.....</b>	<b>42</b>
4.8.1.1 Vezavna mesta napovedana na modelu TLR15_6.....	42
4.8.1.2 Vezavna mesta napovedana na modelu zmodeliranim s strežnikom I-TASSER.....	44
<b>5 RAZPRAVA.....</b>	<b>46</b>
<b>6 SKLEPI.....</b>	<b>52</b>
<b>7 POVZETEK.....</b>	<b>53</b>
<b>8 VIRI.....</b>	<b>55</b>
<b>ZAHVALA</b>	
<b>PRILOGE</b>	

## KAZALO PREGLEDNIC

	str.
Preglednica 1: Strežniki in ponudniki podatkov, ki so na voljo na PMP.....	16
Preglednica 2: Razpoložljiva zaporedja TLR15 v proteinski podatkovni zbirki UniProt....	27
Preglednica 3: Rezultati poravnav s programom BLAST.....	28
Preglednica 4: Pozicije in sekvence domen molekule TLR15.....	30
Preglednica 5: Modeli napovedani iz zaporedja TLR 15, ki so na voljo na spletu.....	32
Preglednica 6: Informacije o šablonah, uporabljenih za modeliranje s strežnikom HHpred .....	33
Preglednica 7: Modeli izračunani s strežnikom SWISS-MODEL.....	34
Preglednica 8: Modeli ter dodatne informacije o modelih izračunanih s strežnikom ModWeb.....	35
Preglednica 9: Informacije o modeliranju s strežnikom Phyre <sup>2</sup> .....	39
Preglednica 10: Šablone in ligandi uporabljeni za napoved vezavnih mest.....	43

## KAZALO SLIK

	str.
Slika 1: Diaciliran in triaciliran lipopeptid (prirejeno po Kang in sod., 2009).....	5
Slika 2: Napovedane domene molekule TLR15 (Matsushima in sod., 2007).....	7
Slika 3: Cone sekvenčne podobnosti in metode za zaznavanje homologije (Venclovas, 2012: 55).....	10
Slika 4: Rezultati poskusov CAMEO iz zadnjih 6 mesecev.....	19
Slika 5: Iskanje obstoječih sekvenc molekule TLR 15 v podatkovni zbirk UniProt.....	20
Slika 6: Vhodna stran programa Clustal Omega za poravnavo več zaporedij.....	20
Slika 7: Vhodna stran algoritma BLAST za poravnavo dveh sekvenc.....	21
Slika 8: Vhodne stran programa WU-BLAST za iskanje po podatkovni bazi TollML.....	22
Slika 9: Vhodna stran PMP.....	23
Slika 10: Interaktivno modeliranje z različnimi strežniki preko PMP.....	23
Slika 11: Domača stran strežnika Phyre <sup>2</sup> .....	24
Slika 12: Pymol - program za grafični prikaz in urejanje molekul.....	25
Slika 13: Domača stran strežnika COACH.....	26
Slika 14: Rezultati poravnave več zaporedij.....	28
Slika 15: Rezultati primerjave sekvence TLR15 s sekvencami v podatkovni bazi TollML	29
Slika 16: Domene in pozicije domen molekule TLR15.....	29
Slika 17: Rezultati iskanja eksperimentalnih struktur in teoretičnih modelov z uporabo sekvence TLR15.....	31
Slika 18: Strukturne informacije modelov s podobnimi sekvencami, ki so na voljo.....	31
Slika 19: Model TLR15 izračunan s strežnikom HHpred.....	33
Slika 20: Modeli izračunani s strežnikom SWISS-MODEL.....	34
Slika 21: Modeli ter dodatne informacije izračunani s strežnikom SWISS-MODEL.....	34
Slika 22: Model TIR domene izračuna s strežnikom ModWeb.....	35
Slika 23: Model TLR15_1 in zaupljivost.....	36
Slika 24: Analiza domen za model TLR15_1.....	36
Slika 25: Pokritost in zaupanje z izbranimi šablonama.....	37
Slika 26: Toplogija modela TLR15_2.....	38
Slika 27: Zaupanje modela TLR15_3 (TLR15_5 in TLR15_6).....	38
Slika 28: Pokritost in zaupanje modela TLR15_4 (levo) ter modelov TLR15_3, TLR15_5, TLR15_6 (desno).....	39
Slika 29: Posamezni LRR-ji domene levcinskih ponovitev.....	40
Slika 30: Primerjava modelov zmodeliranih s Phyre <sup>2</sup> .....	41
Slika 31: Primerjava modelov cele molekule ter modelov domene LRR-jev, zmodeliranih z različnimi strežniki.....	41
Slika 32: Primerjava modelov domene TIR molekule TLR15.....	42
Slika 33: Informacije o napovedi vezavnih mest na modelu TLR15 izračunanim s Phyre <sup>2</sup> 42	42
Slika 34: Kompleksi model (Phyre <sup>2</sup> )-ligand napovedni z metodo COACH.....	43
Slika 35: Informacije o napovedi vezavnih mest na modelu TLR15 izračunanim s protramom I-TASSER.....	44
Slika 36: Kompleksi model (I-TASSER)-ligand napovedni z metodo COACH.....	45

## KAZALO PRILOG

Priloga A: Preglednica šablon uporabljenih za modeliranje strukture molekule TLR

## SLOVARČEK

PDB	Protein Data Bank
TLR	receptorji podobni Toll-u (angl. <i>Toll-like receptors</i> )
TIR	receptor Toll/interlevkin-1
PRR	vzorčno prepoznavni receptor (angl. <i>pattern recognition receptor</i> )
PAMP	s patogenimi mikroorganizmi povezan molekulski vzorec (angl. <i>pathogen-associated molecular pattern</i> )
RNK	ribonukleinska kislina
DNK	deoksiribonukleinska kislina
AK	aminokislina
LRR	levcinska ponovitev (angl. <i>leucin rich repeat</i> )
HCS	ohranjen del levcinske ponovitve (angl. <i>highly conserved segment</i> )
VS	variabilni del levcinske ponovitve (angl. <i>variable segment</i> )
EBI	Evropski bioinformacijski inštitut (angl. <i>The European Bioinformatics Institute</i> )
NCBI	Nacionalni center za biotehnoško informacijo (angl. <i>National Center for Biotechnology Information</i> )
BLAST	Lokalni algoritem za primerjavo sekvenc (angl. <i>Basic Local Alignment Search Tool</i> )
PMP	strežnik Protein Model Portal
(ww)PDB	(Worldwide) Protein Data Bank
HMM	Skriti model Markova
CASP	Critical Assessment of protein Structure Prediction
CAMEO	Continuous Automated Model Evaluation
LRR-NT	N-terminalna levcinska ponovitev
LRR-CT	C-terminalna levcinska ponovitev

## 1 UVOD

Več kot pet desetletij je minilo, odkar so določili prvo tridimenzionalno (3D) strukturo globularnega proteina, mioglobina. Po tem pionirskem delu se je število določenih struktur proteinov izjemno povečalo. Največji arhiv strukturnih podatkov, Protein Data Bank (PDB), trenutno vsebuje več kot 70.000 proteinskih struktur. Ta bogat spekter strukturnih podatkov ponuja edinstveno možnost za boljše razumevanje molekularnih mehanizmov proteinskih funkcij in njihove evolucije (Andreeva, 2012).

Strukture proteinov imajo med drugim tudi ključno vlogo v biomedicinskih raziskavah. Trenutno je za več milijonov proteinov poznano aminokislinsko zaporedje (sekvenca), le majhnemu deležu le-teh pa je eksperimentalno določena struktura. Edini izvedljiv način, da zapolnimo vrzel med sekvencami in strukturami, je računalniško modeliranje (Kryshtafovich in Fidelis, 2009).

Homologno oziroma primerjalno modeliranje temelji na empirični ugotovitvi, da imajo evolucijsko sorodne proteinske domene pogosto podobne 3D strukture. Poleg tega se strukturne lastnosti proteina običajno ohranijo tudi, ko se sekvenca spremeni zaradi mutacij. Zaradi tega je 3D struktura najbolj robustna lastnost homolognih proteinov in je tudi najpogosteje uporabljana za napovedovanje sorodnih proteinov (Venclovas, 2012).

Poseben pomen v proteinski strukturi imajo vezavna mesta za ligande. V zadnjem času se zelo hitro povečuje zanimanje za računalniške metode za napovedovanje in karakterizacijo lokacije vezavnih mest za ligande. Vezavna mesta se lahko prepozna s sočasno kristalizacijo proteina z ligandom, z določitvijo strukturne ali sekvenčne podobnosti znanega vezavnega mesta ali pa z uporabo orodij za napovedovanje vezavnih mest (Laurie in Jackson, 2006).

S pomočjo homolognega modeliranja lahko poleg napovedi struktur, identificiramo aktivna in vezavna mesta proteinov. S pomočjo modelov molekul lahko preučujemo reakcije med proteini, med proteinom in ligandom ter razvijamo inhibitorje. Lahko tudi iščemo, razvijamo ter izboljšujemo ligande za določeno vezavno mesto, napovemo epitope za antigene ter simuliramo interakcije proteinov (Fiser, 2009).

### 1.1 NAMEN DELA

Namen dela je bil s pomočjo homolognega modeliranja napovedati 3D strukturo molekule TLR15 in vezavna mesta za ligande ter preveriti potencialne ligande za protein. S pomočjo bioinformacijskih orodij smo poiskali in primerjali znane sekvence molekule TLR15 ter eno od njih uporabili za modeliranje. Za napoved strukture smo uporabili različne strežnike ter modele primerjali med sabo. Končni model smo analizirali in napovedali vezavna mesta ter potencialne ligande.

## 2 PREGLED OBJAV

### 2.1 KLASIFIKACIJA PROTEINOV

Strukturne podobnosti med proteini se lahko pojavijo na različnih nivojih strukturne organizacije. Podobnosti so lahko lokalne ali globalne. Lokalne so sestavljene samo iz nekaj elementov sekundarne strukture, globalne pa so razširjene čez celo terciarno ali kvartarno strukturo. Te strukturne podobnosti lahko nakazujejo na biološke povezave med proteini in tako omogočajo pomemben vpogled v funkcijo proteinov in strukturno evolucijo (Andreeva, 2012).

V tem poglavju bom opisala osnovne enote strukturne klasifikacije proteinov. Poleg proteinskih domen, ki so najbolj uporabljene, bom predstavila še dodatne enote klasifikacije: motive, ponovitve in proteinske komplekse (Andreeva, 2012).

#### 2.1.1 Proteinske domene

Proteinsko domeno lahko definiramo glede na zvijanje, strukturo, funkcijo ali evolucijo proteinov. V splošnem proteinsko domeno definira en ali več izmed naštetih kriterijev:

- Kompaktna, globularna regija strukture, ki je delno odvisna od ostale polipeptidne verige (strukturna domena); ta regija lahko sestoji iz enega ali več segmentov polipeptidne verige, celotne polipeptidne verige ali več polipeptidnih verig.
- Del proteina, ki se v naravi pojavlja kot samostojna enota ali znotraj več večdomenskih proteinov (evolucijska domena).
- Regija strukture proteina, ki je povezana z določeno funkcijo (funkcijska domena) (Andreeva, 2012).

#### 2.1.2 Ostale klasifikacijske enote

Večina klasifikacij uporablja proteinske domene kot klasifikacijske enote. V okviru klasifikacijske sheme so domene običajno organizirane hierarhično glede na njihove strukturne in evolucijske povezave. Enote, opisane v nadaljevanju, v to hierarhično predstavitev vnesejo dodatno kompleksnost (Andreeva, 2012).

##### 2.1.2.1 Proteinski motivi

Proteinski motiv je lokalna, relativno majhna, neprekinjena regija znotraj polipeptidne verige proteina, ki jo določa dobro definiran nabor lastnosti (strukturnih in/ali funkcijskih).

Obstajata dva tipa motivov:

- sekvenčni – predstavlja ohranjen aminokislinski vzorec, ki je skupen skupini proteinov
- strukturni – kombinacija več sekundarnih elementov s specifično geometrijsko razpolovitvijo. V nasprotju s proteinsko domeno ni kompakten in nima dobro definiranega hidrofobnega jedra (Andreeva, 2012)

##### 2.1.2.2 Proteinske ponovitve

Simetrija in strukturna podvojitev sta široko razširjeni lastnosti naravnih proteinov. Do danes je znanih veliko število proteinskih struktur z notranjo simetrijo in/ali stalno ponavlja-

jočimi se struktturnimi enotami. Te enote se imenujejo proteinske ponovitve in so po navadi tandemsko razporejene v sekvenci in/ali strukturi.

Razlikujemo dva tipa ponovitev:

- sekvenčne – zaporedje ponovitev istega aminokislinskega ostanka ali skupine podobnih aminokislinskih ostankov. Pogosto se enakost sekvenc in število sekvenčnih ponovitev med proteinskimi homologi razlikuje
- Strukturne - katerakoli razporeditev sekundarnih struktturnih elementov, ki se ponovijo znotraj proteinske strukture. Meje sekvenčnih ponovitev pogosto sovpadajo s struktur-nimi ponovitvami, vendar obstajajo tudi izjeme

Proteinske ponovitve se lahko zvijejo v kompaktne domene različnih oblik in kompleksno-sti; pogosto so simetrične. Nekatere homologne ponovitvene strukture se lahko zvijejo in ukrivijo (angl. *bent and coil*) na različne načine, zaradi česar lahko njihova globalna struk-tturna podobnost postane zanemarljiva. Te strukturne različice običajno nastanejo zaradi različnih interakcij med sosednjimi ponovitvami. Iz strukturnih ponovitev lahko nastanejo različne domene (nitaste, globularne, solenoidi in toroidi) (Andreeva, 2012).

Za prepoznavanje ponovitev so na voljo različne metode. Večina metod za identifikacijo sekvenčnih ponovitev uporablja standardne algoritme za primerjavo sekvenc, ki so prilago-jeni za ponovitve. Običajno delajo dobro, kadar je sekvenčna podobnost med ponovitvami velika (ponovitve ne vsebujejo velikih insercij ali delečij) (Andreeva, 2012).

#### 2.1.2.3 Proteinski kompleksi

Večina globularnih in membranskih proteinov se združuje v oligomerne komplekse, ki so sestavljeni iz dveh ali več polipeptidnih verig. Razlikujemo lahko dva tipa oligomernih kompleksov: homomerne in heteromerne. Homomerni so sestavljeni iz dveh identičnih verig, heteromerni pa iz dveh različnih verig. Večina (50-70 %) proteinskih kompleksov je homomernih. V oligomernih kompleksih obstajata dva različna tipa vmesnih površin: izo-logne (homologne) in heterologne. Izologne so oblikovane z enakimi površinami dveh po-denot, medtem ko pri heterolognih površini nista enaki (Andreeva, 2012).

#### 2.1.3 Klasifikacija glede na tipe proteinov

Proteine razdelimo v štiri osnovne skupine z značilnimi sekvencami in struktturnimi la-stnostmi.

- Globularni proteini – so topni v vodnih raztopinah. Običajno se zvijajo v kompak-tne enote, njihova 3D struktura pa odraža interakcijo s topilom. Globularne proteine je relativno enostavno analizirati in kristalizirati, zato so strukturne lastnosti te sku-pine najbolje določene. Tako predstavlja največji delež proteinov z znano strukturo, ki so na voljo za modeliranje
- Nitasti proteini – ta skupina vsebuje proteine, kot so kolagen, keratin, elastin, itd. ki so večinoma netopni. Glede na sekundarno strukturo jih lahko razdelimo v tri pod-skupine: trojna vijačnica, niti beta-listov in alfa-nitasti proteini
- Membranski proteini – večina integriranih membranskih proteinov sestoji iz trans-membranskih  $\alpha$ -vijačnic, ki so običajno organizirane v sveženj. Klasifikacija

membranskih proteinov primarno temelji na njihovih tipičnih arhitekturnih in topoloških lastnostih

- Intrinzično nestrukturirani proteini - regije proteinov ali tudi celi proteini, ki v naravnih oblikah nimajo urejene strukture, v funkcionalnem stanju pa lahko prehajajo iz neurejenega v urejeno stanje. Najbolj pomembno je, da te regije oz. proteini niso v skladu z osnovnim načelom strukturne biologije, ki pravi, da proteinske sekvene določajo edinstveno 3D strukturo proteina, ta pa določa njegovo funkcijo (Andreeva, 2012)

## 2.2 RECEPTORJI PODOBNI TOLL-U

Receptorji podobni Toll-u oziroma receptorji TLR so membranski proteini, ki sprožijo prirojen imunski odziv po tem ko se nanje vežejo različne spojine patogenih organizmov (Botos in sod., 2011).

Beseda Toll je nemškega izvora in v prevodu pomeni nekaj fantastičnega, odločnega, genialnega. V znanstvenem kontekstu sta jo prvič uporabila Nusslein-Volhard in Anderson. Z njim sta poimenovala gen, ki sta ga odkrila z genetskim pregledovanjem *Drosophila* in ima pomembno vlogo pri razvoju insektov (Gay in Gangloff, 2007).

Sekvenca gena Toll, ki so jo določili leta 1988 je razkrila, da je protein sestavljen iz treh delov: N-terminalne regije (domene), transmembranske regije ter C-terminalne regije. V zgodnjih devetdesetih letih so odkrili, da je C-terminalna domena zelo sorodna receptorju za interlevkin-1 (IL-1R), ki je prisoten tudi pri vretenčarjih. Takrat so odkrili tudi, da je ta domena poleg razvoja insektov vpletena tudi v nastanek vnetnega odziva, ki ga vršijo človeške imunske celice. Po tem pomembnem odkritju so pri vretenčarjih identificirali 10 receptorjev, ki so bili homologni Toll-u in so jih poimenovali kot receptorji podobni Toll-u (Gay and Gangloff, 2007).

Leta 1994 so okarakterizirali rastlinski protein (N protein), ki omogoča odpornost rastline na virus tobačnega mozaika. C-terminalna domena tega proteina je podobna citoplazemski domeni Toll-a in IL-1R, kar nakazuje, da je ta ohranjena domena vpletena v obrambo pred pateogenimi organizmi pri dveh različnih kraljestvih, živalskem in rastlinskem. Domeno so poimenovali receptor Toll/interlevkin-1 oziroma TIR (O'Neil in sod., 2013).

### 2.2.1 Vloga v imunskemu sistemu

Pri začetni fazi okužbe prirojeni imunski sistem sproži hiter vnetni odziv, ki zavira razširjanje kužnega agensa. Pri vretenčarjih temu odgovoru sledi pridobljena imunost, pri kateri specifični celični receptorji B in T prepoznajo tuje molekule in vzpodbudijo odgovor, s katerim odpravijo patogene organizme. Antigeni receptorji prirojenega imunskega sistema so vnaprej določeni (angl. *germline encoded*) ter evolucijsko izbrani da prepoznajo patogene spojine, ki so osnovne za preživetje telesu tujih organizmov ali endogene molekule, ki jih gostitelj sprošča kot odgovor na okužbo (Botos in sod., 2011).

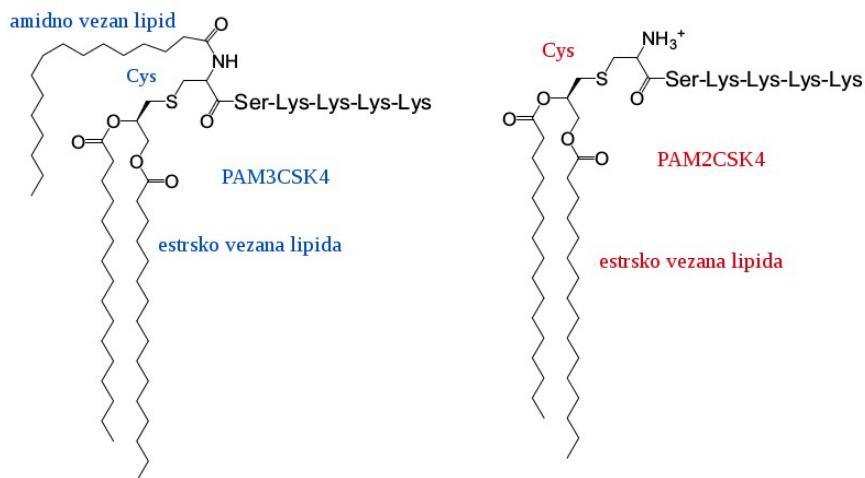
Prepoznavanje prirojenega imunskega sistema temelji na zaznavanju molekularnih struktur, ki so unikatne za mikroorganizme. Posredovano je z receptorji prirojenega imunskega sistema, ki se imenujejo vzorčno prepoznavni receptorji (angl. pattern recognition receptors oziroma PRRs). Posebnost vzorčnega prepoznavanja je, da ima receptor PRR široko specifičnost ter veže veliko število molekul, ki imajo skupen strukturni motiv ali vzorec. Tarče PRR-jev se imenujejo »s patogenimi mikroorganizmi povezani molekulski vzorci«

(angl. pathogen-associated molecular patterns oziroma PAMPs). Za njih je značilno, da so med mikroorganizmi določenega razreda nespremenljivi. So strukture ali produkti metabolnih poti, ki so edinstvene za mikroorganizme (s tem je možno razlikovanje med lastnimi in tujimi molekulami) ter imajo ključno vlogo v fiziologiji mikroorganizma. Pri bakterijah in glivah so PAMP-i običajno komponente celične stene, pri virusih pa nukleinske kisline (Medzhitov, 2007).

Obstaja več različnih razredov PRR-jev, najbolj okarakterizirani med njimi so TLR-ji (Medzhitov, 2007). PRR-je so identificirali v serumu, na celični površini, v endosomih in v citoplazmi (Botos in sod., 2011).

#### 2.2.1.1 Tarče (ligandi) TLR-jev (PAMP-i, ki jih prepozna TLR-ji)

- Bakterijski: lipopolisahardi gram negativnih bakterij, peptidoglikani gram pozitivnih bakterij, lipoarabinomanan mikobakterij, diacilirani in triacilirani lipopeptidi (slika 1) bakterij mikobakterij in mikoplazem, flagelin bičastih bakterij ter bakterijska genomska deoksiribonukleinska kislina (DNK) bogata z nemetiliranimi CpG otočki (Kumar in sod., 2009)



Slika 1: Diaciliran in triaciliran lipopeptid (prirejeno po Kang in sod., 2009)

- Virusni: dvoverižna ribonukleinska kislina (RNK), enoverižna RNK, nemetilirani CpG otočki DNK, virusni glikoproteini in v nekaterih primerih tudi proteini, ki niso del strukture virusa in se nahajajo v izvenceličnem prostoru (Lester in Li, 2013)
- Glivni: fosfolipomanani,  $\beta$ -glukani, glukoronoksilosilomanani (Kumar in sod., 2009)
- Praživalski: glikoinozitolfosfolipidi (GIPL), glikozilfosfatidilinozitol (GPI), nenasiličeni alkilacilglicerol in lipofosfoglikan (LPG), genomska DNK ter profilinu podobni proteini (Kumar in sod., 2009)

#### 2.2.2 Struktorna biologija TLR-jev

TLR-ji so membranski receptorji tipa I in so sestavljeni iz treh domen: N-terminalne domene, ki se nahaja izven celice ter prepoznavajo ligande, transmembranske domene – heliks ter C-terminalne, signalne domene, ki se nahaja znotraj celice (Botos in sod., 2011).

### 2.2.2.1 N-terminalna domena

N-terminalna domena ali ektodomena molekule TLR je glikoprotein dolg 550-800 aminokislin (AK). Nahaja se izven celice ali v endosomih, kjer se srečuje in prepozna tuje molekule. Sestavljena je iz tandemskih kopije motivov – levcinskih ponovitev (LRR) (Botos in sod., 2007).

Za levcinske ponovitve je značilno da vsebujejo z levcinom bogate ponavljače se sekvenčne vzorce in so dolge 20-30 AK. Njihova glavna lastnost je 11 aminokislinskih ostankov dolga sekvenca z značilnim zaporedjem. Proteini in domene, ki vsebujejo tandemске ponovitve dveh ali več LRR-jev, oblikujejo čedalje bogatejšo proteinsko superdružino LRR-jev (Bella in sod., 2008).

Vsi LRR-ji so sestavljeni iz zelo ohranjenega dela (angl. *highly conserved segment* oziroma HCS) ter variabilnega dela (angl. *variable segment* oziroma VS). HCS sestoji iz 11 ali 12 aminokislinskih ostankov z naslednjim zaporedjem: L<sub>xx</sub>L<sub>x</sub>L<sub>xx</sub>N(Cx)xL, kjer L predstavlja levcin, izolevcin, valin ali fenilalanin, ki oblikujejo hidrofobno jedro. N predstavlja asparagin, treonin, serin ali cistein (C predstavlja cistein, serin ali asparagin) ter x predstavlja katerokoli AK. Variabilni del se razlikuje tako po dolžini kot zaporedju AK (Gong, 2010). Glede na sekvenčne analize proteinov LRR naj bi obstajalo vsaj sedem različnih poddržin: ribonukleaznemu inhibitorju podobni LRR-ji (angl. *RI-like*), proteinu SDS22+ podobni LRR-ji (angl. *SDS22-like*), cistein vsebujoča poddržina LRR-jev, bakterijski tip in značilni tip (angl. *typical type*), za rastline specifični tip ter TpLRR poddržina LRR-jev (Kobe in Kajava, 2001). Pri molekulah TLR sta prisotna značilni tip (xxLxxxxLxxLxx) in bakterijski tip (xxLPx(x)LPxx). Vsi LRR-ji pri molekulah TLR imajo N- in C-terminalni LRR, ki imata neobičajno zaporedje in se ne ujemata z nobenim tipom (Gong in sod., 2010).

3D struktura vseh LRR-jev je enaka in predstavlja strukturo zanke. Ko se te sestavijo v protein tvorijo solenoidno strukturo pri kateri so hidrofobni ostanki usmerjeni v notranjost. Ti tvorijo stabilno jedro,  $\beta$ -trakovi pa se poravnajo in oblikujejo z vodikovimi vezmi povezano  $\beta$ -ploskev. Ker so  $\beta$ -trakovi bolj blizu zapakirani, kot tisti deli LRR-jev, ki ne vsebujejo  $\beta$ -struktur, ima solenoid zakriviljeno obliko. Rezultat je struktura, ki vsebuje konkavno in konveksno površino (Botos in sod., 2011). Konkavna stran LRR domene je definirana s paralelnimi  $\beta$ -ploskvami in v splošnem vsebuje mesta za vezavo ligandov. Konveksna stran je prepletena z različnimi strukturnimi elementi:  $\alpha$ -vijačnice,  $3_{10}$  vijačnice, vijačnice poliprolina tipa II,  $\beta$ -zavoji in tudi kratki  $\beta$ -trakovi (Bella in sod., 2008). Poleg konveksne in konkavne površine ima struktura LRR še dvigajoč se lateralno/bočno površino, ki sestoji iz zank, ki povezujejo  $\beta$ -trakove s konveksno površino ter spuščajoč se lateralno/bočno površino na druge strani (Botos in sod., 2011).

Ektodomene molekule TLR običajno vsebujejo 19-25 LRR-jev, skupno zaporedje ene LRR molekule TLR pa je dolgo približno 24 aminokislinskih ostankov (Botos in sod., 2011).

### 2.2.2.2 Transmembranska domena

Razen hidrofobnosti za te segmente molekul ni značilen poseben ohranjen vzorec, imeli pa naj bi ključno vlogo pri aktivaciji receptorja. Ektodomena TLR-jev je povezana s citoplazemsko domeno TIR z eno transmembransko  $\alpha$ -vijačnico (Gay and Gangloff, 2007), za ka-

tero je značilno da je zgrajena iz približno 20, večinoma hidrofobnih aminokislinskih ostankov (Botos in sod., 2011).

### 2.2.2.3 Domena Toll/interlevkin-1 receptor

Domena je dobila ime po signalni domeni družine receptorjev interlevkin-1, s katerimi izkazuje homologijo. Homologi domene TIR se nahajajo tudi v nekaterih rastlinskih proteinih, ki so odgovorni za odpornost, kar nakazuje da je ta star motiv služil za imunost že pred ločitvijo rastlin in živali (Botos in sod., 2011).

Glede na določene strukture domen TIR molekul TLR1, TLR2, TLR10, IL-1RAPL in MyD88 je le-ta sestavljena iz paralelnih  $\beta$ -plošč iz petih trakov ( $\beta A - \beta E$ ) ter petih vijačnic ( $\alpha A - \alpha E$ ), ki so povezane s površinsko izpostavljenimi zankami. Ohranjenost sekvence med domenami TIR je 20-30 %, predvsem pomembna pa naj bi bila ohranjenost BB zanke, ki povezuje  $\beta B$ -plast z  $\alpha B$ -vijačnico in je odgovorna za prenos signala pri TLR molekul. TIR domena je sestavljena iz 135 do 160 AK ostankov (Fekonja in sod., 2012).

### 2.2.3 Toll-u podobni receptorji pri ptičih

Imunski odgovor pri ptičih in sesalcih je precej podoben, čeprav sta se rodova ločila že pred približno 300 milijoni let. Znanje o imunologiji ptičev se je povečalo s pridobitvijo genomske sekvene kokoši (lat. *Gallus gallus*). Na podlagi doslej znanih raziskav naj bi obstajalo 10 ptičjih molekul TLR (1, 2, 3, 4, 5, 6, 7, 15 in 21). Molekula TLR15 naj bi bila edinstvena za ptiče, čeprav je filogenetsko sorodna skupini molekul TLR2 (Brownlie in Allan, 2010)

#### 2.2.3.1 TLR15

S sekvenčno analizo levcinskih ponovitev pri vretenčarjih so se ukvarjali že Matsushima in sod. (2007). S pomočjo programa PFAM, so določili LRR-je molekul TLR različnih organizmov. Za molekulo TLR15 so napovedali, da ima poleg N- in C-terminalnega še 20 LRR-jev, od katerih je LRR3 zaradi svoje dolžine nespecifičen.

cTLR15	LRR12	LLFLDLSSNRI SMLPDLPI
SIGNAL MRILIGSLYFYFISFLFSKVNG	LRR13	LQQLDISNSDI KIIPPRFKSLSN
LLRRNT FLTQRTSPVSSFPFYNSYLNLSVSQAQAKPT	LRR14	VTVFNIQNNKL TEMHPEYFPST
LxxLxLxxNxL	LRR15	LTTCDISKNKL KVLSLTKALEN
LRR1 ARALNFSYNAI EKITKRDFFEGFH	LRR16	LESLNVSGNLI TRLEPACQLPS
LRR2 LEVLDLSHNHI KDIEPGAFENLLS	LRR17	LTNLDSSHNLI SELPDHLLGQSLLM
LRR3 LVSVDSLSDK NLLVSGLAPHLKLIPPTGASGSPSQIYMYFQ	LRR18	LKHFNLSGNKI SFLQRGSLPAS
KSAEAALEPSAPAELLPHLEDPPNPNGNVNP	LRR19	LEELDISDNAI TTIVQDTFGQLTS
RFRQRRTTEENKTSPPAATLRPDLCGAPI	LRR20	LSVLTVQGKHF FCNCNDLYWFVNII
LRR4 NGLLDLSRTKL SNEELTAKLDADLCQAQLGT	LRRCT	IRNPHLQINGKDDLRCSPFPDRRGSLVKSSNLTLHCSLG
LRR5 VLEFNISHSDL EMDLLSLFILFLPMKD		IQMAITAC
LRR6 IQSVDASYNRI TINNIDVEAICHFPFSN	TRANS	MAILVVVLVTGLCW
LRR7 FSFLNISNNPI NSLETVCCLPAS	CYTOP	RFDGLWYVRMGWYWCMAKRQYKKRPEPKPDFAFISYSEH
LRR8 ITVIDLSFTNI STIPANFAKKLSK		DADWTKEHLLKKLETDGFKICYHERDFKPGHPVGLNIFYC
LRR9 LERMYVQGNQL IYTIVRPENPSATPRPPPCTVQ		IENSHKVLVLPSPVNSCWCQEELYFAEHRVLDENQDSL I
LRR10 ISAISLVRNQA GTPIESLPES		MVVLEDLPPDSVPQKFSKLRKLLKRKTYLKWSPEEHKQK I
LRR11 VKHLKVSNCSI VELPEWFANRMQE		FWHQLAAVLTTNEPLVRAENGPNEDVIEME

Slika 2: Napovedane domene molekule TLR15 (Matsushima in sod., 2007)

Poleg levcinskih ponovitev so napovedali tudi sekvenco signalnega peptida, transmembranske domene ter citoplazemskega domene (slika 2).

TLR15 se molekulsko razlikuje od ostalih znanih molekul TLR in se ne izraža tkivno specifično, zaradi česar je težko določit potencialne ligande. Nerren in sod. (2010) so identificirali povečano izražanje mRNA za molekulo TLR15 pri piščancih, kot odziv na že znane ligande (agoniste) molekul TLR in različne vrste bakterij. Kot znane agoniste TLR so uporabili flagelin (agonist molekule TLR5), lipopolisaharid (agonist molekule TLR4), peptido-glikan, triaciliran lipopeptid, lipotehoično kislino (agonisti molekule TLR2) ter nemetilirane CpG otočke (agonist molekule TLR9). Za izražanje molekule TLR15 so uporabili naslednje bakterije: *Salmonella enterica*, *Escherichia coli*, *Enterococcus gallinarum* ter *Rhodococcus equi*. Študije so pokazale, da se izražanje TLR15 znatno poveča ob stimulaciji s vsemi bakterijami razen z bakterijo *Rhodococcus equi*, medtem ko se izražanje ne poveča ob stimulaciji z agonisti. Rezultati študije torej nakazujejo, da posamezni agonisti molekul TLR niso ligandi molekule TLR 15 ter da bi naj le-ta prepoznavala edinstvene termostaabilne komponente po Gramu+ in po Gramu- bakterij, ki pri kokoših povzročajo bolezen.

Oveni in sod. (2013) so v študiji eksperimentalno pokazali, da diaciliran lipopeptid iz bakterije *Mycoplasma synoviae* v kokošjih celicah sproži povečano izražanje kokošjega receptorja TLR15 iz česar so sklepali, da je diaciliran lipopeptid ligand kokošjega receptorja TLR15.

### 2.3 BIOINFORMATIKA

Pojem »bioinformatika« naj bi se prvič uporabil v sredini 80 let prejšnjega stoletja in naj bi opisoval uporabo informacijske znanosti in tehnologije v biologiji. Definicije so se skozi leta spremajale, danes pa naj bi bila najbolj pravilna naslednja: »Bioinformatika je znanost o tem, kako informacija v bioloških sistemih nastane, se prenaša, sprejema ter interpretira« (Ramsden, 2009).

Tehnološki napredek na področju sekveniranja DNK, proteomike in določevanja struktur je vodil do velike količine genomskeh in proteomskeh podatkov, podatkov pridobljenih s pomočjo mikročipov ter funkcionalno genomiko (Lopez in sod., 2002). Področje bioinformatike se je razvilo prav zaradi potreb naravoslovcev po analizi in interpretaciji velike količine podatkov (Cohen, 2004). V zelo kratkem času je bioinformatika postala zelo aktivno raziskovalno področje in je zelo interdisciplinarna. Zahteva vsaj matematično, biološko, fizikalno in kemijsko znanje, njena uporaba pa lahko v prihodnosti zahteva še znanje iz računalništva, kemijskega inženirstva, biotehnologije, medicine, farmakologije, itd (Ramsden, 2009).

#### 2.3.1 Homologno modeliranje

Homologno modeliranje se uporablja za konstruiranje 3D strurnih modelov proteinov (tarča) na podlagi poravnave aminokislinske sekvence tarče s sorodnim proteinom z znano strukturo (šablona). Homologno modeliranje sestoji iz 4 korakov: (1) identifikacija sorodnega proteina, ki ima eksperimentalno določeno strukturo in se zato lahko uporabi kot strurna šablona, (2) kartiranje (angl. *mapping*) aminokislinskih ostankov sekvence tarče na strukturo šablone - sekvenčno-strurna poravnava, (3) generiranje 3D modela tarčnega proteina na osnovi sekvenčno-strurne poravnave in (4) ocenjevanje pravilnosti dobljenega modela. Celoten proces lahko ponavljamo, dokler ocenjena kvaliteta modela ni zadovoljiva ali modela ne moremo izboljšati (Venclovas, 2012).

Pred skoraj dvema desetletjema je na internetu postal dostopen prvi avtomatizirani strežnik za modeliranje – SWISS-MODEL. Od takrat se jih je razvilo še nekaj: ModWeb, Robetta,

HHpred, I-TASSER, Pcons, Phyre ali M4T. Osnovni koraki pri primerjalnem struktturnem modeliranju so izbira šablone, poravnava šablone s tarčo, oblikovanje modela in ocena kvalitete modela. Glavni cilj za avtomatizacijo osnovnih korakov je narediti te tehnologije dosegljive tistim, ki niso strokovnjaki na področju bioinformatike. Torej, lažja uporaba računalniških orodij, ki drugače zahteva specializirano tehnično znanje, vzdrževanje programske opreme za modeliranje ter upravljanje z veliko količino sekvenčnih in struktturnih podatkov (Bordoli in Schwede, 2012).

Validacija kvalitete modelov z uporabo ocenjevalne funkcije je zelo pomembna, saj kvaliteta modelov določa njihovo uporabo za določene aplikacije v znanosti. Prva ocena za pričakovano kvaliteto strukturnega modela je identiteta sekvenc med tarčo in šablono. Kvaliteta modela pa ni nujno odvisna od identitete, saj lahko kljub majhni podobnosti dobimo pravilne modele (Bordoli in Schwede, 2012).

Osnovne metode za primerjavo proteinskih sekvenc so bile predstavljene pre približno 40 leti. Začetni poskusi meritev podobnosti proteinskih sekvenc so temeljili na globalni poravnavi sekvenc, kjer je bil vključen vsak AK ostanek. Ker so pa manj sorodni proteini bili podobni le na določenih regijah oziroma v bližini aktivnega mesta, so se začele razvijati tudi metode za lokalno poravnavo (Altschul, 1991). Poleg metod za globalne in lokalne poravnave obstajajo še metode za poravnavo več sekvenc, v vsakem primeru pa so poravnane ovrednotene z uporabo utežne matrike (sheme), ki poda oceno o podobnosti med sekvenca-m oziroma sekvencami. (Henikoff in Henikoff, 1992).

### 2.3.2 Metode za sekvenčno-struktурno poravnavo

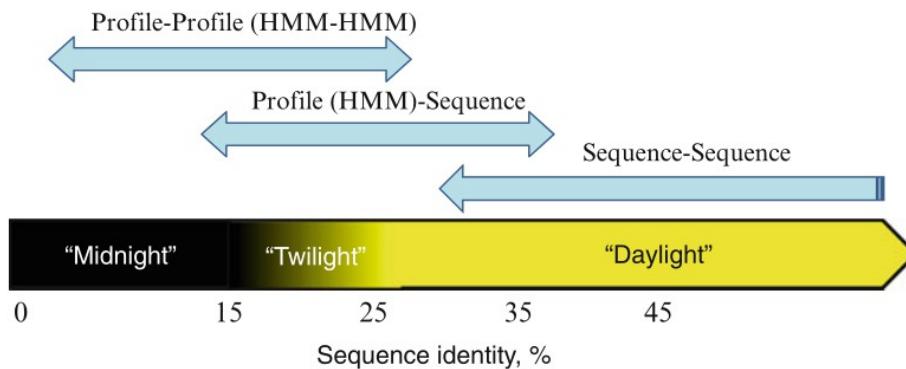
Ko identificiramo primeren struktturni homolog (šablono), je najpomembnejši dejavnik za kvaliteto dobljenega modela pravilna poravnavna tarče in šablone (sekvenčno-strukturna poravnava). Če ju prekrijemo, dobimo za strukturno podobne regije ujemanje med ostanki. Cilj sekvenčno-strukturne poravnave je to ujemanje čim bolj pravilno reproducirati. Če tarča in šabloni nista zelo sorodni, lahko obstajajo regije z velikimi struktturnimi razlikami. Te struktturno različne regije so največkrat rezultat insercij, delecij ali večjih sprememb v aminokislinski sekvenci. V takih regijah določitev ujemanja ostankov ni vedno očitna in je včasih tudi nesmiselna. Število eksperimentalno določenih struktur konstantno narašča, zato lahko tarča pogosto poravnamo z večjim številom struktturnih šablon, ki so na voljo. Običajno ne moremo doseči pravilne poravnave po celotni dolžini sekvence z eno samo šabloni, zato lahko različne regije tarče poravnamo z različnimi šablonami. To nam omogoča, da izboljšamo model, vendar hkrati dodatno oteži postopek modeliranja (Venclovas, 2012).

Ker na začetku poznamo le aminokislinsko zaporedje tarče, je sekvenčna primerjava glavno sredstvo za identifikacijo sorodnih proteinov (šablon) z znanimi eksperimentalno določenimi 3D strukturami (Venclovas, 2012).

Če je podobnost poravnanih sekvenc statistično pomembna (podobnost, ki po vsej verjetnosti ni naključna), jih obravnavamo kot sekvence s skupnim evolucijskim izvorom. Nadalje to pomeni, da so verjetno podobne tudi njihove 3D strukture. Glede na evolucijsko oddaljenost med proteinimi potrebujemo za določevanje odnosa med njimi metode različnih kompleksnosti. Te metode lahko razdelimo v skupine glede na naraščajočo kompleksnost informacije o sekvenci, ki jo uporabljajo (slika 3):

1. poravnavi dveh sekvenc (sekvenčno-sekvenčna poravnavi),

2. profilno-sekvenčne poravnave in HMM-sekvenčne poravnave (profilno-sekvenčna poravnava z uporabo skritih modelov Markova),
3. profilno-profilne poravnave in HMM-HMM poravnave (Venclovas, 2012).



**Slika 3:** Cone sekvenčne podobnosti in metode za zaznavanje homologije (Venclovas, 2012: 55)

#### 2.3.2.1 Sekvenčno-sekvenčne metode

Najzgodnejše metode za zaznavanje homologije uporabljajo poravnavo dveh sekvenc in so konceptualno najenostavnejše. Uporabljajo le sekvence dveh proteinov, matriko zamenjav aminokislinskih ostankov in algoritem za poravnavo. Sekvenčno-sekvenčno metode običajno vrnejo podatke o statistični pomembnosti dobljenih poravnav, ki jih lahko uporabimo za iskanje po bazah sekvenc. BLAST je nedvomno najbolj znano in priljubljeno orodje za iskanje po bazah, ki temelji na poravnavi dveh sekvenc. Poleg BLAST-a se pogosto uporablja tudi FASTA in Ssearch. Z naštetimi programi lahko hitro dobimo začetno oceno o težavnosti homolognega modeliranja. Primerni so za zaznavanje evolucijsko sorodnih proteinov, ki imajo enakih več kot 25-30 % ostankov – to območje sekvenčne podobnosti imenujemo tudi »dnevna« cona (angl. *daylight zone*). Tudi v teh primerih poravnave večinoma potrebujejo izboljšave. Pravilnost poravnave lahko pričakujemo, če se poravnani sekvenci ujemata v več kot 40-50 % aminokislinskih ostankov in poravnava vsebuje le malo ali nič vrzeli (Venclovas, 2012).

Uporabnost sekvenčno-sekvenčnih primerjav za odkrivanje sekvenčno-struktturnih poravnав je omejena, zato se te metode čedalje manj uporabljajo neposredno za modeliranje. Kljub temu je to začetni korak v praktično vseh bolj sofisticiranih metodah sekvenčne primerjave, ki uporabljajo informacije iz več sorodnih sekvenc. Izboljšave v tej začetni primerjavi imajo lahko zato velik vpliv na končni rezultat (Venclovas, 2012).

#### 2.3.2.2 Profilno-sekvenčne in HMM-sekvenčne metode za poravnavo

Pri večji evolucijski oddaljenosti – sekvenčna podobnost v coni »someraka« (angl. *twilight zone*) – sekvenčno-sekvenčna poravnava ne zadošča za zanesljivo identifikacijo homologov. V takih primerih so bolj učinkovite metode, ki uporabljajo sekvenčne profile ali skrite modele Markova (Hidden Markov Models – HMMs). Pri obeh pristopih uporabimo poravnavo skupine sorodnih sekvenc za izgradnjo celovitega statističnega modela, iz katerega so razvidna mesta, kjer lahko najverjetneje pride do insercij ali delecij. Primerjava profila s sekvencami iz podatkovne baze nam tako omogoča bolj občutljivo zaznavanje homologov

in generiranje pravilnejših poravnava. Trenutno je najbolj uporabljen metoda PSI-BLAST (angl *position-specific iterated BLAST*).

Najboljši metodi za HMM-sekvenčno poravnavo sta HMMER in SAM. HMM-ji delujejo podobno kot sekvenčni profili, s tem da uporabljajo verjetnostne metode za določitev ocenjevalnih parametrov. HMM-ji upoštevajo tudi verjetnosti insercije in delecije na vsakem mestu profila, s čimer poskušajo bolje predstaviti lastnosti evolucije proteinske sekvence. Te verjetnosti so znotraj proteinske sekvence seveda zelo odvisne od pozicije, zaradi različnih strukturnih in/ali funkcionalnih omejitev. Insercije/delecije so lahko škodljive znotraj strukturnega jedra, medtem ko so bolje tolerirane v strukturno variabilnih regijah, izpostavljenih topilu. Tako kot sekvenčni profili tudi HMM-ji obravnavajo posamezno mesto neodvisno od ostalih in tako ne morejo zajeti više razrednih medsebojnih odvisnosti, ki lahko obstajajo v proteinskih sekvenkah. Kljub metodološkim prednostim metod, ki temeljijo na HMM-jih, se te še vedno ne uporabljajo tako pogosto kot PSI-BLAST. En od razlogov za to je, da so občutno počasnejše, poleg tega pa je težko oblikovati iterativen postopek, ki bi deloval tako dobro kot pri PSI-BLAST. Kljub temu je področje HMM precej napredovalo. Program HMMER se lahko primerja z BLAST-om po hitrosti iskanja, njegova iterativna različica Jackhmmer pa dosega rezultate, po občutljivosti in pravilnosti poravnave, primerljive s PSI-BLAST (Venclovas, 2012).

#### 2.3.2.3 Profilno-profilne in HMM-HMM metode za poravnavo

Evolucijske odnose, ki so preveč oddaljeni, da bi jih zaznali s sekvenčno-sekvenčno ali profilno-sekvenčno (HMM-sekvenčno) primerjavo (cona »polnoči«), lahko v določenih primerih vseeno zaznamo z metodami, ki temeljijo na profilno-profilnih ali HMM-HMM poravnovah. Te metode dodajo nivo kompleksnosti tako, da primerjajo dva sekvenčna profila (HMM-ja) namesto profila (HMM-ja) s posamezno sekvenco. V tem primeru nas ne zanima pripadnost sekvence določeni družini, ampak evolucijska povezanost dveh sekvenčnih družin. Ta pristop nam je omogočil še bolj občutljivo detekcijo homologov in izboljšanje pravilnosti poravnav. Najbolj znani predstavniki metod HMM-HMM so HHsearch in PRC, najbolj znani predstavniki profilno-profilnih metod za primerjavo pa COMPASS, COMA in PROCAIN. Zaenkrat sta obe metodologiji v aktivnem razvoju in še ni jasno, katera se bo izkazala za bolj uspešno, saj ima vsaka svoje prednosti in slabosti (Venclovas, 2012).

### 2.3.3 Metode za poravnavo več sekvenc

Metode za poravnavo več sekvenc (angl. *multi sequence alignment* ozziroma MSA) za razliko od ostalih niso narejene za zaznavanje homolognih sekvenč, ampak za poravnavo množice homolognih sekvenč, ki smo jih že identificirali s katero od prej opisanih metod. Metode MSA selahko uporabljajo: (1) za izboljšanje kvalitete MSA-jev, iz katerih so narejeni profili (HMM-ji) za iskanje homologije in poravnavo in (2) če sta tako tarča kot šablona v množici sekvenč za poravnavo, lahko poravnavo (tarče in šablone) pridobimo neposredno iz dobljenih MSA (Venclovas, 2012).

Metode MSA iz množice sekvenč ustvarijo poravnavo, v kateri stolpci predstavljajo evolucijsko (strukturno) ekvivalentne ostanke. Večina modernih orodij za MSA uporablja hevriističen pristop imenovan progresivna poravnavna (angl. *progressive alignment*). Pri tej strategiji se na podlagi podobnosti med pari sekvenč najprej zgradi približno vodilno drevo (angl. *guide tree*) za poravnavo. S pomočjo tega drevesa se najprej poravnajo najbolj soro-

dne sekvene. Te poravnave so zatem poravnane med sabo, dokler niso vse sekvene vključene v MSA. ClustalW je eden prvih predstavnikov metod za progresivno poravnavo in je še vedno med najbolj priljubljenimi programi. Glavna pomanjkljivost te strategije je, da se napake med gradnjo drevesa in začetnimi koraki parnih poravnav razširijo na celotno poravnavo. Zato lahko ClustalW dobro poravna le sorodne sekvene, medtem ko so poravnave raznolikih množic sekvenc lahko slabe. Obstaja več pristopov, s katerimi se lahko izognemo problemom, povezanih z uporabo progresivne poravnave (Venclovas, 2012).

En od teh je uporaba iterativnega izpopolnjevanja (angl. *iterative refinement*). Zaenkrat obstajata dve metodi, ki uporablja to strategijo: MAFT in MUSCLE. Obe sta zelo hitri in prilagodljivi. Druga strategija za izboljšanje progresivne poravnave je uporaba informacije o skladnosti. Taki metodi sta: T-coffee in ProbCons. Načeloma so metode, ki temeljijo na skladnosti, bolj pravilne od tistih, ki temeljijo na iterativnem izpopolnjevanju, ampak so tudi bolj računsko zahtevne. Ostale strategije za izboljšavo pravilnosti poravnave uporabljajo kombinacijo različnih metod (npr. M-coffee), ali pa vključujejo dodatne informacije. Te so lahko evolucijske (npr. dodatne homologne sekvene) ali strukturne (ker se 3D struktura spreminja počasneje kot sekvena). Eden od nedavno razvityih programov, PROMALS, uporablja številne vire dodatnih informacij. Nekateri programi, kot so PROMALS3D (naslednik PROMALS) in 3DCoffee/Expreso, avtomatsko vključijo informacije o 3D strukturah, ki so na voljo (Venclovas, 2012).

Potrebno je še poudariti, da je izbira optimalne metode MSA odvisna od situacije. V primeru razmeroma enakih sekvenc (čez 35 % sekvenčne identitete oz. »dnevna« cona), bo verjetno katerakoli metoda naredila pravilno poravnavo. Pravilnost poravnave se začne zniževati, ko sekvenčna podobnost pada v cono »somraka« (<25 %) in/ali kadar obstaja le majhno število sekvenc. V teh primerih so kljub počasnosti bolj primerne metode, ki uporabljajo dodatne sekvene in/ali informacije (Venclovas, 2012).

### 2.3.4 Hibridne metode, popolno integrirani avtomatski strežniki in meta-strežniki

Vse večje število sodobnih metod za modeliranje uporablja kombinacijo več sekvenčnih in strukturnih lastnosti za izpeljavo sekvenčno-strukturne poravnave. Poleg tega se za izračun proteinskih modelov, ki temeljijo na homologiji, običajno istočasno obravnavajo številne poravnave z več šablonami ali njihovimi fragmenti (Venclovas, 2012).

Druga priljubljena možnost je uporaba meta-pristopov, ki uporabljajo kombinacijo rezultatov različnih algoritmov za identifikacijo najbližje strukturne šablone in najpravilnejše sekvenčno-strukturne poravnave. Trenutno najbolj popularne metode, ki so se izkazale za uspešne na CASP-u (Critical Assessment of protein Structure Prediction), so: I-TASSER, pro-Sp3-TASSER, SAM-T08, Phyre, MULTICOM, pGen THREADE, GenSilico in Pcons.net sta dva meta-strežnika, ki se neprestano razvijata in posodabljata (Venclovas, 2012).

Čeprav obstaja veliko število popolnoma samodejnih metod za homologno modeliranje, je vseeno potrebno upoštevati, da uporaba bolj sofisticiranih postopkov ne zagotavlja boljše kakovosti končnega modela. Vedno znova ugotavljamo, da tehnika, s katero pridemo do končnega modela, ni tako pomembna kot optimalna izbira šablon in izboljšave sekvenčno-strukturne poravnave. Zato lahko metode, ki tvorijo pravilnejše poravnave, včasih presežejo bolj kompleksne metode (Venclovas, 2012).

## 2.4 BIOINFORMACIJSKA ORODJA

### 2.4.1 Evropski bioinformacijski inštitut

Evropski bioinformacijski inštitut (EBI), del Evropskega laboratorija za molekularno biologijo, že od leta 1997 omogoča dostop do prevladujočih bioinformacijskih podatkovnih baz in orodij (Mcwilliam in sod., 2009). Ogromna količina bioloških podatkov je najpomembnejši del aktivnosti inštituta in njihova razpoložljivost zahteva identifikacijo standarnih metod za njihov dostop, kajti le tako se lahko omogoči obsiren in enoten pogled za vse večje število uporabnikov (Lopez in sod., 2002).

S podatkovnimi bazami lahko dostopamo do sekvenc nukleotidov in proteinov, z orodji pa lahko iščemo homologije in podobnosti med sekvencami, določimo proteinske funkcije, analiziramo strukture proteinov, napovemo strukture proteinov ter poravnamo več sekvenc hkrati (Lopez in sod., 2003).

Spletne storitve Evropskega bioinformacijskega inštituta so naslednje (Mcwilliam in sod., 2009):

- pridobivanje podatkov (WSDbfetch, ChEBI, WS, Integr8 WS, API, MartService, EB-eye, ArrayExpres, IntAct, SRS, QuickGO, UniProt, JAPI),
- orodja za analizo (InterProScan, EMBOSS, CENSOR, Phobius, Soaplab),
- iskanje podobnosti (Fasta, WU-BLAST, NCBI-BLAST, PSI-BLAST, PHI-BLAST, PSI-Search, MPSRCH, SCANPS),
- poravnava več sekvenc (ClustalW, Clustal Omega, Kalign, MAFFT, MUSCLE, T-Coffee),
- za analizo struktur (DaliLite, MaxSprout, MSDFold (SSM)),
- literatura in ontologija (CiteXplore, Whatizit, OLS, WSSBO, MIRIAM, PICR, Bio-Models).

#### 2.4.1.1 Universal Protein Resource (UniProt)

UniProt je podatkovna baza, za katero skrbi konzorcij, ki sestoji iz skupin iz Evropskega bioinformacijskega inštituta (EBI), Švicarskega inštituta za bioinformatiko (Swiss Institute of Bioinformatics - SIB) in javnega bioinformacijskega vira genomskeh in proteomskeh raziskav iz Univerze v Georgetownu (Protein Information Resource – PIR). Omogoča izčrpen, visoko kvaliteten in prosto dostopen vir informacij o proteinskih sekvencah in funkcijski anotaciji. Z organizacijo biološkega znanja olajša znanstveno raziskovanje ter omogoča raziskovalcem boljše razumevanje tudi bolj zahtevnejših področij biologije (UniProt Consortium, 2014).

Uniprot vključuje 4 podatkovne baze:

- UniProt Knowledgebase (UniProtKB), ki je sestavljena iz dveh delov: UniProtKB/Swiss-Prot (preverjen del, ki vsebuje ročno anotirane vnose, z informacijami iz literature in računalniških analiz), UniProtKB/TrEMBL (nepreverjen del z avtomatsko anotiranimi vnosi) (UniProt Consortium, 2014)

- UniProt Archive (UniParc) - obsežen arhiv vseh proteinskih sekvenc iz glavnih javno dostopnih proteinskih baz (UniProt Consortium, 2014)
- UniProt Reference Clusters (UniRef) – v tej bazi so združene sekvence iz UniProt-KB ter UniParc podatkovnih baz. Baza zajema vse proteinske sekvence, katerih število pa se zaradi gručenja enakih sekvenc in fragmentov sekvenc precej zmanjša (ni presežnosti). To omogoča hitrejše iskanje podobnih sekvenc (Suzek in sod., 2007)
- UniProt Metagenomic and Environmental Sequence (UniMES) – omogoče prost dostop do genomskeh informacij iz okoljskih vzorcev (metagenomskih podatkov) (UniProt Consortium, 2008)

#### 2.4.1.2 Clustal Omega

Je najnovejši program iz družine Clustal, s katerim lahko poravnamo več sekvenc. Je pravilnejši v primerjavi z ostalimi hitrimi metodami, ki so v uporabi, pravilnost poravnave pa se lahko primerja tudi z ostalimi počasnejšimi metodami. Njegova posebnost je, da uporabniku omogoča ponovno uporabo poravnave, vsakič ko so na voljo nove sekvence, brez ponovnega poravnavanja. Program ima tudi možnost, kjer lahko uporabnik doda sekvence že obstoječi poravnavi ali pa uporabi obstoječe poravnave (Sievers in sod., 2011).

### 2.4.2 Nacionalni center za biotehnoško informacijo

Nacionalni center za biotehnoško informacijo (NCBI) je del Nacionalne Medicinske knjižnice Združenih držav Amerike, ki deluje pod okriljem Nacionalnega inštituta za zdravje v ZDA. S pomočjo računalniških pristopov se ukvarja z osnovnimi vprašanji na področju biologije in omogoča dostop do biomedicinskih informacij ter bioinformacijskih orodij (Hoeppner in Ostell, 2013).

Viri informacij na strani zajemajo naslednja področja: vodič po spletni strani ter Entrez sistem s podatkovnimi bazami literature (PubMed, PubMed Central, NLM Catalog, Medical Subject Headings, NCBI Bookshelf), taksonomije, podatkovne baze DNK in RNK (RefSeq, GenBank, PopSet, Sequence Read Archive, Trace Archive, BioSample), proteinske podatkovne baze (RefSeq, GenBank, Protein Clusters, HIV-1/Human Protein Interaction Database), orodja za analizo sekvenc (BLAST, BLAST podatkovna baza, Genomic BLAST, Primer BLAST, COBALT), viri informacij o genih in njihovem izražanju (Gene, RefSeqGene, The Conserved CDS Database, Gene Expression Omnibus, UniGene, HomoloGene, Probe, Biosystems), viri informacij o genomih (BioProject, Gernome Reference Consortium, Clone Database, Epigenomics, Influenza Genome Resource), viri informacij o genetiki in medicini (dbGap, dbVar, dbSNP, OMIA, dbMHC, dbLRC, dbRBC), viri informacij o kemikalijah in bioloških testih (PubChem), viri informacij o domenah in strukturah (Molecular Modeling Database, Conserved Domain Database) (NCBI Resource Coordinators, 2014).

#### 2.4.2.1 BLAST

BLAST (Basic Local Alignment Search Tool) je algoritem za lokalno primerjavo sekvenc. Je heveristična metoda, ki je sposobna zaznati šibke ampak biološko pomembne sekvenčne podobnosti in je hitrejša v primerjavi z ostalimi heverističnimi metodami (Altschul, 1990).

Pomembna izboljšava je bila dosežena z razvojem programa CS-BLAST (angl. *context-specific*). Ta metodološka inovacija je poleg višje občutljivosti homologne detekcije dose-

gla tudi pomembne izboljšave kvalitete poravnave (Biegert in Söding, 2009). CS-BLAST je še posebej obetaven za uporabo pri sekvenkah brez zaznanih homologov (angl. *singleton sequences*). Pomanjkanje sorodnih sekvenc namreč izključuje uporabo profilno-sekvenčnih in profilno-profilnih metod, opisanih v naslednjih dveh razdelkih (Venclovas, 2012).

Primerjava profila s sekvencami iz podatkovne baze nam omogoča bolj občutljivo zaznavanje homologov in generiranje pravilnejših poravnav. Trenutno je najbolj uporabljen metoda PSI-BLAST (angl. *position-specific iterated BLAST*). Ta s poravnavo najboljših zadetkov prvega kroga iskanja z BLAST-om zgradi pozicijsko-specifično ocenjevalno matriko (angl. *position-specific scoring matrix – PSSM*), ki nadomesti generično matriko zamenjav (npr. BLOSUM ali PAM) v naslednjem krogu iskanja BLAST. Postopek lahko ponovimo večkrat in zmeraj uporabimo najboljše nove sekvence za prilagoditev profila. Z večanjem števila ponovitev v profil vključujemo čedalje manj sorodne sekvene, pri tem pa profil ostaja značilen za sekvenčno družino. PSI-BLAST je zato zelo močno orodje za iskanje in primerjavo sekvenc, ki lahko pogosto zazna in poravna homologe s sekvenčno identiteto tudi manjšo od 15 % (v coni »somraka« in »polnoči«). Ker njegov osnovni korak temelji na BLAST-u, tudi PSI-BLAST obravnava sosednja mesta v sekvenci neodvisno. Tako kot CS-BLAST tudi CSI-BLAST (angl. *context-specific iterated BLAST*) dosega boljše rezultate od PSI-BLAST, kar nakazuje, da je vključitev sekvenčnega konteksta v sekvenčne ali profilne primerjave obetavna možnost za izboljšave (Venclovas, 2012).

Od leta 1995 je na voljo program WU-BLAST, ki omogoča hitro, fleksibilno in zanesljivo metodo za iskanje podobnosti po bioloških podatkovnih bazah. Je prvi program te vrste, ki je omogočal presledkovne poravnave s statistično pomembnimi ocenami. Z algoritmom WU-BLAST lahko iščemo po 12 proteinskih podatkovnih bazah ter 34 nukleotidnih podatkovnih bazah. Številni spletni strežniki uporabljajo BLAST algoritem za iskanje sekvenčne podobnosti. Eden takih je tudi TollML opisan v naslednjem razdelku (Lopez in sod., 2003).

#### 2.4.3 Podatkovna baza TollML

TollML je podatkovna baza, ki vsebuje vse sekvene TLR iz proteinske podatkovne baze NCBI. Vnosi so razdeljeni glede na družine TLR (TLR1-TLR23) in potem še dodatno razdeljeni na tri nivoje glede na kategorije struktturnih motivov (Gond in sod., 2010):

- signalni peptid, ektodomena, transmembranska domena in Toll/IL-1 receptor domena
- LRR-ji posamezne ektodomene
- visoko ohranjeni deli, variabilni deli ter insercije posameznega LRR-ja

Anotacija posameznega vnosa TLR vsebuje (Gond in sod., 2010):

- informacije o upravljanju podatov: TollML ID in datum spremembe ali dostopa
- primarni podatki iz podatkovne zbirke (NCBI): zaporedje FASTA, biološka definicija, celična informacija, glikozilacija in ligandi
- klasifikacija proteinske družine
- povezave do ostalih podatkovnih zbirk: NCBI, Swiss-Prot, PDB, LRRML, PubMed
- Informacija o motivih

TollML omogoča tudi iskanje podobnosti po podatkovni bazi z orodjem WU-BLAST (opisan v prejšnjem razdelku). Uporabnik lahko po želji omeji iskanje in išče podobnosti le znotraj določene družine TLR (Gond in sod., 2010)

#### 2.4.4 Strežnik Protein Model Portal

Strežnik Protein Model Portal (PMP) je bil razvit z namenom, da spodbudi uporabo 3D molekulskih modelov v biomedicinskih raziskavah, tako da omogoča priročen in obširen dostop do informacij o strukturah proteinov. Edina podatkovna baza, ki vsebuje informacije o strukturah proteinov, je wwPDB (Worldwide Protein Data Bank), ki pa vsebuje le eksperimentalno določene strukture, medtem ko PMP poleg teh vsebuje tudi teoretične modele, izračunane z metodami za modeliranje (Haas in sod., 2013).

Pri iskanju po PMP, izvirajo podatki o eksperimentalnih strukturah iz najnovejše zbirke PDB, primerjalni modeli pa iz skladišča že prej določenih modelov. Dejstvo je, da za vse danes znane sekvence proteinov ne moremo izračunati modelov in da so za vsak model lahko s časom na voljo boljše šablone. PMP nam omogoča tudi, da sočasno zaženemo modeliranje z več trenutno najsodobnejšimi viri za modeliranje, s katerimi lahko pridobimo najnovejše modele. Uporaba različnih neodvisnih metod nam lahko pokaže, za katere dele strukturnega modela lahko pričakujemo, da so bolj ali manj zanesljivi. Regije proteinov, ki so z različnimi neodvisnimi metodami napovedane enako so verjetneje pravilnejše. PMP omogoča tudi oceno kvalitete modela z različnimi orodji (Bordoli in Schwede, 2012).

V preglednici 1 so navedeni strežniki in ponudniki podatkov, ki so trenutno na voljo na PMP.

**Preglednica 1:** Strežniki in ponudniki podatkov, ki so na voljo na PMP

Strežniki za modeliranje	HHpred, I-TASSER, ModWeb, M4T, SWISS-MODEL Workspace
Strežniki za oceno kvalitete	ModEval, ModFOLD3, QMEAN
Ponudniki proteinskih modelov	PSI-Structural-Genomics Centers, ModBase, SWISS-MODEL Repository, GPCRDB

##### 2.4.4.1 HHpred

HHpred je strežnik za detekcijo homologije proteinov ter napovedovanje strukture in je prvi, ki je primerjal pare profilov skritih modelov Markova. Omogoča iskanje po več bazah, kot so PDB, SCOP, PFAM, SMART, COG in CDD. Kot vhodna datoteka se lahko uporabi posamezna sekvencia ali pa poravnava več sekvenc. V nekaj minutah dobimo rezultate v uporabniku prijaznem formatu. Možnosti za iskanje vključujejo lokalne ali globalne poravnave ter primerjavo sekundarnih struktur (angl. *scoring secondary structure similarity*). HHpred lahko poravna dve ali več sekvenci (z eno ali več šablonami) ter iz teh poravnav s pomočjo programa MODELLER izračuna 3D strukturni model (Söding in sod., 2005).

##### 2.4.4.2 SWISS-MODEL

Najnovejša verzija strežnika se imenuje SWIS-MODEL Workspace in je prosto dostopno spletno delovno okolje, kjer lahko uporabnik izračunava in shranjuje rezultate raznih računalniških nalog za oblikovanje homolognih modelov. SWIS-MODEL Workspace nudi dostop do programske opreme in baz, potrebnih za izvedbo štiri osnovnih korakov homolognega modeliranja. Za način izvedbe modeliranja je na voljo »Automated« način, kjer sta

aminokislinsko zaporedje oziroma akcesijska številka proteina dovolj za izračun struktурnega modela. V primerih, kjer sta tarča in sekvenca evolucijsko precej oddaljeni je priporočljiv način »Alignment«, kjer se za izračun modela uporabijo poravnave več sekvens (tarče, šablone in sorodnih proteinov). Strežnik ponuja tudi »Project« način, ki omogoča uporabniku, da sam preuči in obdela poravnavo (med tarčo in šablono) ter oblikuje model na podlagi poravnave, ki jo sam določi (Bordoli in Schwede, 2012).

Programi, kot so SWISS-MODEL, generirajo strukturne koordinate modela na podlagi ostankov tarče in ustreznih aminokislin strukture šablone. Regije proteinov, kjer ne moremo dobiti informacij od šablone (insercije in delekcije v viačnicah), se običajno oblikujejo s pomočjo knjižnic proteinskih fragmentov (angl. *Protein backbone fragment libraries*). Na koncu sledi validacija kvalitete celotnega modela z uporabo orodij za ocenjevanje, kot so ANOELA ali QMEAN. SWISS-MODEL Workspace ponuja tudi dodatna orodja, kot so programi za anotacijo funkcij in domen, identifikacijo šablone ter oceno strukture (Bordoli in Schwede, 2012).

#### 2.4.4.3 ModWeb

ModWeb je strežnik za avtomatsko primerjalno modeliranje proteinskih struktur. Kot vhodno datoteko lahko uporabimo eno ali več sekvens v formatu FASTA. Strežnik izračuna model na podlagi šablonskih struktur, ki so na voljo v PDB. ModWeb je odvisen od ModPipe, ki je popolnoma avtomatizirana programska oprema za primerjalno modeliranje in lahko izračuna primerjalne modele za veliko število proteinskih zaporedij z uporabo različnih šablon ter sekvenčno-strukturnih poravnav. Sekvenčno-strukturne poravnave so pridobljene na dva načina; s poravnavo profila PSI-BLAST tarčne sekvence s vsako sekvenco šablon, pridobljenih iz PDB zbirke, ter s pregledovanjem tarčne sekvence po bazi profilov z uporabo programa IMPALA, ki primerja posamezno sekvenco s podatkovno bazo PSSM-jev, ki jih generira PSI-BLAST (Schäffer in sod., 1999). Pomembne poravnave, ki pokrivajo različne regije tarčnega zaporedja so izbrane za modeliranje. Modeli, so izračunani za vsak sekvenčno-strukturni zaditek z uporabo programa MODELLER. Dobljeni modeli se nato ocenijo z različnimi kriteriji za ocenjevanje kvalitete (Eswar in sod., 2003).

#### 2.4.4.4 Phyre<sup>2</sup>

Phyre<sup>2</sup> je strežnik za napovedovanje proteinske strukture. Obstajata dva načina modeliranja: »Normal« in »Intensive«. Pri prvem načinu strežnik naredi potencialne 3D modele proteina, ki temeljijo na poravnavi z znanimi proteinskimi strukturami.

Proces modeliranja v načinu »Normal« vključuje:

- zaznavanje sekvenčnih homologov s programom Psi-Blast
- napoved sekundarne strukture s programoma Psi-pred in Diso-pred
- izdelavo skritega modela Markova (HMM) tarčnega zaporedja, ki temelji na homologih zaznanimi s Psi-blast
- pregled HMM-ja po knjižnici HMM-jev proteinov, ki imajo eksperimentalno določene strukture (te knjižnice se posodablja vsak teden), ter poravnava le-teh
- izračun 3D modelov proteina na osnovi poravnave
- Modeliranje insercij in delecij ter modeliranje stranskih verig AK

- Napoved transmembranskih vijačnic ter topologije s programom memsat-svm

Nalčin »Intensive« je identičen načinu »Normal« do točke izgradnje modela. Pri obeh načinih dobimo enake rezultate, vendar so koraki po izračunu modelov pri načinu »Intensive« sledeči:

- izbira šablon z največjim zaupanjem, ki pokrivajo največji del tarčne sekvence
- končni model je izračunan s programom Poing na podlagi vseh modelov; *ab initio* so zmodelirana mesta, kjer je homologija premajhna da bi lahko poiskali šablone

Modeliranje v načinu »Intensive« je trenutno omejeno na manj kot 1000 aminokislinskih ostankov (Kelly in Jefferys, 2011).

## 2.5 METODE ZA PREPOZNAVANJE VEZAVNIH MEST ZA LIGANDE NA PROTEI- NIH

V zadnjem času se zelo hitro povečuje zanimanje za računalniške metode za zaznavanje in karakterizacijo lokacije, kjer se vežejo ligandi. Vezavna mesta se lahko prepozna s sočasno kristalizacijo proteina z ligandom, z določitvijo strukturne ali sekvenčne podobnosti znanega vezavnega mesta ali pa z uporabo orodij za napovedovanje vezavnih mest (Laurie in Jackson, 2006).

Ligandi se običajno vežejo na specifična mesta tarčnega proteina. Ta vezavna mesta naj bi se zaradi posebnih lastnosti razlikovala od ostalih delov proteina. Te lastnosti so lahko elektrostatski potencial, velikost votline na površini proteina, itd (Xie in Hwang, 2012).

Veliko metod za izračun vezanih mest/žepov, ki temeljijo na geometriji je bilo razvitih z radi domnevanja, da se ligand vedno veže v največjo votlino v tarčnem proteinu. Ta domneva naj bi se izkazala za pravilno v 70 %. Kljub vsemu, natančna določitev praznih prostorov (votlin) na proteinski površini ter njihova veličina nista trivialni, kajti različni algoritmi ne dajo vedno istih rezultatov za isti protein (Xie in Hwang, 2012).

Drugačen tip metode temelji na izračunu energije oziroma energetsko najbolj ugodnih regij za vezavo ligandov. Te metode razvrščajo sonde (običajno sfere, ki predstavljajo majhne molekule ali kemijsko skupino z določeno funkcijo, npr. hidrofobnostjo) okrog tarčnega proteina in računajo interakcijsko energijo med sondom in atomi tarčnega proteina s katerimi so v stiku. Sledi gručenje sond z ugodnimi interakcijskimi energijami, ki so identificirane kot energetsko najbolj ugodna regija, na podlagi katerih se napove vezavno mesto za ligand. Ker te metode ne identificirajo največje votline, so bolj uspešne v primerih kjer geometrijske metode niso uspešne (kjer se ligand veže v majhne votline). V nasprotnem primeru, izračuni energije lahko izpustijo največje votline, metode so dražje ter lahko zahtevajo sofisticirane sheme gručenja. Slabost teh metod je tudi, da ligandi običajno niso poznani in je zato težko oblikovati sondu (Xie in Hwang, 2012).

Ker je na voljo vse več proteinskih struktur, so čedalje bolj pomembne metode, ki temeljijo na homologiji. Te so lahko bolj zanesljive, vendar zahtevajo homologne strukture z znanimi vezavnimi mesti za ligande (Xie in Hwang, 2012). Pri metodah, ki uporabljajo strukture za napoved vezavnih mest prevladujeta dva pristopa. Pri prvem je žep za vezavo liganda določen s pomočjo prepoznavanja površinskih votlin 3D strukturnih modelov tarčnega proteina. Gre za *ab initio* modeliranje vezavnih mest, ker so napovedi izvedene brez uporabe šablon. Slabost je ta, da lahko dobimo veliko lažno pozitivnih rezultatov. Drug pristop je pridobitev informacij o vezavnem mestu s pomočjo že znanega proteina (šablone), ki ima

podobno globalno in/ali lokalno strukturo. Glede na CASP je ta metoda daleč najbolj pravilna, predvsem za tarče z bližnimi homologi, ki imajo v podatkovne baze vnesene komplekse ligand-protein (Yang in sod., 2013).

Druga strategija, ki si pomaga s strukturami, je statistično določevanje značilnosti, ki naj bi se evolucijsko ohranile na vezavnih mestih (angl. *propensity based methods*). Tam, kjer je vezavno mesto, naj bi bila sestava aminokislin precej drugačna od ostale površine proteina (Xie in Hwang, 2012). Na podoben način delujejo tudi metode, ki temeljijo na podobnosti sekvenč in se zanašajo na domnevo, da so aminokislinski ostanki, ki so del vezavnega mesta funkcionalno pomembni in se zato skozi evolucijo ohranjajo. Prednost te metode je, da lahko dobimo napoved iz same sekvence, vendar je natančnost napovedi slaba, kajti veliko ostankov ki niso del vezavnih mest je tudi lahko ohranjenih, ker imajo druge pomembne funkcije (Yang in sod., 2013).

Ker imajo različne metode različne prednosti in slabosti, imajo večji potencial metode, ki združujejo različne strategije, npr. Fpocket, MetaPocket ter metoda COACH. Slednji uporablja rezultate različnih metod za meta-napoved, ki je trenutno najbolj uspešna med vsemi (Xie in Hwang, 2012).

### 2.5.1 Metoda COACH

COACH je prosti dostopni spletni meta-strežnik, ki zajema napovedi petih programov: TM-SITE, S-SITE, COFACTOR, FINDSITE in ConCavity (Yang in sod., 2013).

TM-SITE je algoritem, ki napove vezavna mesta za ligande s pomočjo struktурno podobnih proteinov z znanimi vezavnimi mesti (šablonami). Algoritem S-SITE za razliko od TM-SITE za napoved vezavnih mest uporablja primerjavo sekvenčnih profilov, s katerimi išče in prepoznavajo evolucijsko ohranjena vezavna mesta. Kombinacija teh dveh metod z ostalimi »state-of-the-art« metodami, lahko poveča učinkovitost napovedi za 1/6 (Yang in sod., 2013).

Uspešnost metode COACH se testira tudi s poskusi CAMEO (Continuous Automated Model Evaluation) poskusi. CAMEO je bil oblikovan z namenom, da vseskozi ocenjuje računalniške metode in je komplementaren poskusom CASP. Prednost poskusov CAMEO pred poskusi CASP je, da ocenjuje metode kontinuirano ter na podlagi večjega števila tarč. COACH sodeluje v poskusih CAMEO od decembra 2012 (Yang in sod., 2013).

Na sliki 4 so prikazani rezultati poskusov CAMEO zadnjih 6 mesecev, kjer je COACH na prvem mestu.

▼ 6-months - [2013-11-15 - 2014-05-16]							Targets		
Server Name	Avg. response time (hh:mm:ss)	#Predictions	Targets			Average Accuracy		Server Status	Server Access
			#Submitted	#Modeled	Modeled (%)	All targets	Modeled targets		
COACH	13:14:42	1063	1104	1063	96.3	0.84	0.86	Active	public
RaptorX-Binding	19:08:01	884	1104	884	80.1	0.72	0.77	Active	public
Naive Pocket	02:21:13	991	1104	991	89.8	0.71	0.73	Active	public
HHfuncs	04:36:23	665	1104	665	60.2	0.67	0.78	Active	public
FunFOLD2-FN	35:14:22	823	1104	823	74.5	0.64	0.68	Active	public
Naive Conservation	00:21:28	1062	1104	1062	96.2	0.63	0.64	Active	public
Naive Homology	01:34:50	277	1104	277	25.1	0.59	0.85	Active	public
FunFOLD	47:00:46	399	807	399	49.4	0.58	0.67	Deactivated on 2014-03-25	public
FunFOLD2-LB	37:02:02	411	1104	411	37.2	0.56	0.66	Active	public

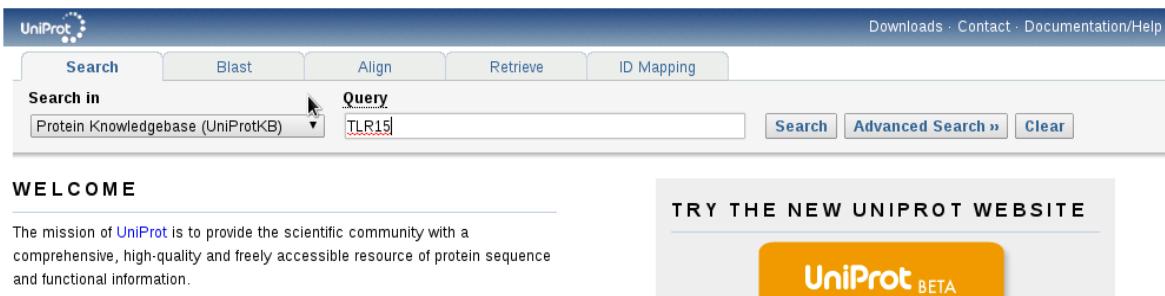
Slika 4: Rezultati poskusov CAMEO iz zadnjih 6 mesecev

### 3 METODE

#### 3.1 ANALIZA OBSTOJEČIH ZAPOREDIJ MOLEKULE TLR15 S POMOČJO EVROPSKEGA BIOINFORMACIJSKEGA INŠTITUTA

##### 3.1.1 Iskanje ter pridobitev obstoječih zaporedij s podatkovno bazo UniProt

S pomočjo iskalnika v podatkovni zbirkki UniProt smo pridobili vse obstoječe sekvene molekule TLR15. Iskalnik nam ponudi možnost iskanja po različnih bazah. Naše nastavite so bile naslednje: v okence za vnos iskanja smo vnesli »TLR15« ter izbrali podatkovno bazo UniProtKB (Protein Knowledgebase) (slika 5).

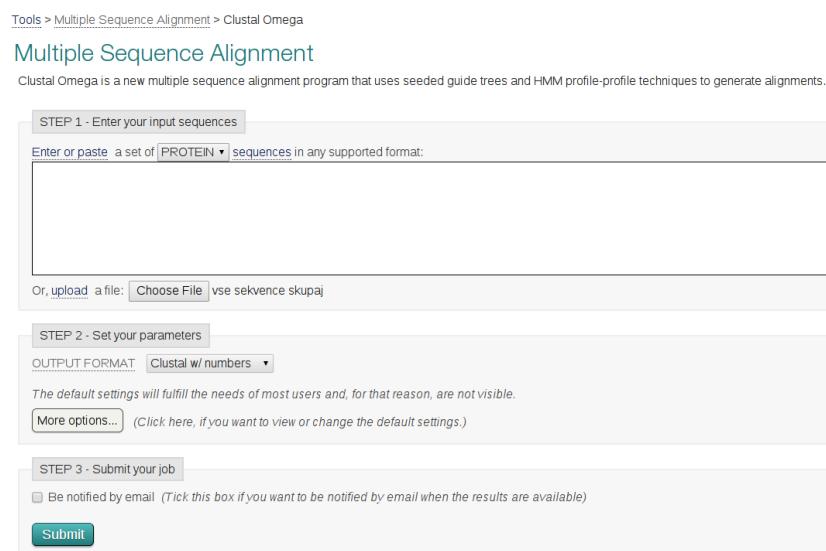


Slika 5: Iskanje obstoječih sekvenc molekule TLR 15 v podatkovni zbirkki UniProt

Izbrali smo eno sekvenco posamezne vrste ptiča (13 vrst), jo izvozili v formatu FASTA ter jo shranili kot datoteko .txt (naredili smo 13 datotek). Naredili smo tudi datoteko .txt, ki vsebuje format FASTA vseh 13 zaporedij skupaj.

##### 3.1.2 Poravnava več sekvenč s programom Clustal Omega

Datoteko .txt, ki vsebuje vse sekvence smo uporabili za poravnavo s programom Clustal Omega. S tem programom smo poravnali in primerjali podobnost sekvenč vseh 13 organizmov naenkrat. Vhodne nastavite za poravnavo so bile privzete in so prikazane na sliki 6.



Slika 6: Vhodna stran programa Clustal Omega za poravnavo več zaporedij

### 3.2 ANALIZA PODOBNOSTI MED ZAPOREDJI MOLEKULE TLR15 S POMOČJO NACIONALNEGA CENTRA ZA BIOTEHNOLOŠKO INFORMACIJO

Za iskanje proteinske sekvene TLR15 ter primerjavo zadetkov smo uporabili proteinsko podatkovno bazo ter algoritmom BLAST, oba na voljo na strežniku Nacionalnega centra za biotehnološko informacijo.

#### 3.2.1 Poravnavo sekvenc z algoritmom BLAST

Algoritmom BLAST smo uporabili za poravnavo dveh proteinskih sekvenc (opcija »Align two or more sequences«). Kot vhodne datoteke, smo uporabili datoteke .txt s sekvencami posameznih organizmov, ki smo jih predhodno izvozili v format FASTA iz proteinske podatkovne zbirke UniProt. Kot sonda oziroma tarčno sekvenco (angl. *query sequence*) smo uporabili sekvenco organizma *Gallus gallus*, ki so ga posekvenirali najprej in ima največ zadetkov v bazi UniProt. S to sekvenco smo poravnali sekvene TLR15 ostalih 12 organizmov (angl. *subject sequence*), ki so bili razpoložljivi v zbirki UniProt (slika 7).

The screenshot shows the 'Align Sequences Protein BLAST' page of the NCBI BLASTp suite. At the top, there are tabs for 'blastn', 'blastp' (which is selected), 'blastx', 'tblastn', and 'tblastx'. Below the tabs, there are two main sections: 'Enter Query Sequence' and 'Enter Subject Sequence'. Each section has a text input field, a 'Choose File' button, and a note indicating 'No file chosen'. There are also 'From' and 'To' fields for specifying a subject subrange. Under each section, there is a 'Job Title' input field and a note to enter a descriptive title. A checked checkbox labeled 'Align two or more sequences' is present. At the bottom of the page, there is a 'Program Selection' section where 'blastp (protein-protein BLAST)' is selected. A large blue 'BLAST' button is at the bottom left, and a note says 'Search protein sequence using Blastp (protein-protein BLAST)'. There is also an option to 'Show results in a new window'.

Slika 7: Vhodna stran algoritma BLAST za poravnavo dveh sekvenc

### 3.3 ANALIZA MOLEKULE TLR15 Z ORODJEM TOLLML

#### 3.3.1 Primerjava našega zaporedja z zaporedji v podatkovni zbirki TollML z algoritmom WU-BLAST

Izbrano sekvenco smo primerjali z vsemi sekvencami v podatkovni bazi TollML. Vhodne nastavitev so prikazane na sliki 8. V polje za iskanje smo vnesli tarčno sekvenco molekule TLR15, izbrali pBLAST (protein BLAST) metodo, podatkovno bazo TLR (kar pomeni, da algoritrom išče med vsemi družinami TLR), ocenjevalno matriko PAM70 ter število poravnav 10.

## BLAST (BLAST-WU)

TollML ID :

Protein Name:  no name

AA Sequence: (20 amino acids in one-letter code)  
 MRILIGSLYFYFISFLFSKVNGFLTQRTSPVSSFPFYNSYLNLSVSQAQAPKTARALNFSYNAIEKITKRDSEGHVLEVLDLSHNHDKDIEPGAFENLLSLSVSDLSFN  
PHLKLIPSGASGPSQIYMFQKSAEAALEPSAPAELLPHLEDPPNPNGVNPRFRQRRTEE  
NKTSPPAATLRPDLCGAPINGLLDSRTKLSNEELTAKLDADLCQAQLGTLEFNIHSSDL

BLAST method:

Database:

Scoring matrix:

Number of alignments (max 200):  10

Slika 8: Vhodne stran programa WU-BLAST za iskanje po podatkovni bazi TollML

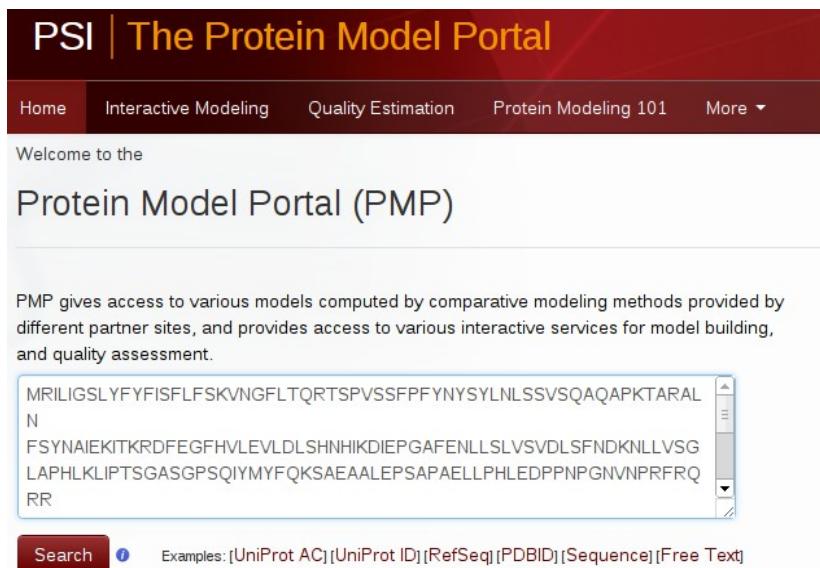
### 3.4 ISKANJE EKSPERIMENTALNIH STRUKTUR IN TEORETIČNIH MODELOV TER INTERAKTIVNO MODELIRANJE S POMOČJO STREŽNIKA PROTEIN MODEL PORTAL

Glede na razpoložljivost podatkov in primerjavo sekvenc s Clustal Omega in BLAST, smo se odločili da bomo 3D strukturo molekule TLR15 napovedovali s proteinsko sekvenco kočki (*Gallus gallus*). Sekvenco smo pridobili iz proteinske baze podatkov UniProt. V nadaljevanju je format FASTA sekvence molekule TLR15.

```
>tr|C4PBP0|C4PBP0_CHICK TLR15 OS=Gallus gallus GN=TLR15 PE=2 SV=1
MRILIGSLYFYFISFLFSKVNGFLTQRTSPVSSFPFYNSYLNLSVSQAQAPKTARALNFSYNAIEKITKRDSEGHVLEVLDLSHNHDKDIEPGAFENLLSLSVSDLSFN
DKNLLVSGLAPHKLIPTSGASGPSQIYMFQKSAEAALEPSAPAELLPHLEDPPNPNGVNPRFRQRRTEE
PGVNPRFRQRRTENKTSPPAATLRPDLCGAPINGLLDSRTKLSNEELTAKLDADLCQAQLGTLEFNIHSSDL
DLCQAQLGTVLEFNISHSDLEMDLLSLFILFLPMKDIQSVDASYNRITINNIDVEAI
CHFPFSNFSFLNISNNPINSLETVCLPASITVIDLSFTNISTIPANFAKKLSKLER
MYVQGNQLIYTVRPENPSATPRPPPGTVQISAISLVRNQAGTPIESLPESVKHL
KVSNCISVELPEWFANRMQEELLFLDLSSNRISMLPDPLPLSQLQQLDISNSDIKIIPPR
FKSLSNLTVFNIQNNKLEMHPEYFPSTLTCDISKNLKVLSLTKALENLESLN
VSGNLITRLEPACQLPSLTNLDSSHNLISELPDHLGQSLLMLKHFNLSGNKISFLQR
GS LPASLEELDISDNAITTIVQDTFGQLTSLSVLTQGKHFNCNDLYWFVNHYIRN
PHLQINGKDDLRCSPPDRRGSLVKSSNLTLHCSLGQMAITACMAILVVVLVTGL
CWRFDGLWYVRMGWYWCMAKRRQYKKRPNKPFDASFISYSEHDADWTKEHLL
KKLETDGFKICYHERDFKPGHPVLGNIFYCIENSHKVLVLSPSFVNSCWCQYELY
FAEHRVLDENQDSLIMVVLEDLPPDSVPQKFSKLRKLLKRKTYLKWSPEEHKQKI
FWHQLAAVLKTTNEPLVRAENGPNEDVIEME
```

#### 3.4.1 Iskanje eksperimentalnih struktur in teoretičnih modelov

Na domačo stran portala smo kot vhodno datoteko vnesli sekvenco TLR15 (z dolžino 868 aminokislin) organizma *Gallus gallus* (C4PBP0) v formatu FASTA (slika 9).



Slika 9: Vhodna stran PMP

### 3.4.2 Iskanje modelov s podobnimi sekvencami

Glede na to, da se naša sekvanca ni ujemala z nobenim od modelov, ter ni pripadala nobeni od eksperimentalno določenih struktur, smo s pomočjo portala zagnali algoritem BLAST. To opcijo nam je program, s vsemi privzetimi nastavitevami, ponudil za iskanje modelov, ki imajo podobne sekvence. V primeru, da se zaporedje ne ujema z nobenim eksperimentalnim modelom, nam portal sam ponudi opcijo za interaktivno modeliranje (slika 10).

### 3.4.3 Interaktivno modeliranje

PMP nam omogoča da zaženemo proces modeliranja molekule ter ocenjevanje modela iz različnih strežnikov naenkrat. Na portalu smo posredno modeliranje zagnali dvakrat. Prvič

The screenshot shows the "Interactive Modeling" section of the PMP homepage. It lists several modeling services with their status and results:

- SWISS-MODEL**: Status: request submitted, Result: [Swiss-Model Website](#)
- I-TASSER**: Status: request failed, Failed submission: Due to the limit of computer resource, I-TASSER server is only devoted to academic users. Please use your institutional email address and submit your sequence again. Thanks!
- ModWeb**: Status: request submitted, Result: [ModWeb Website](#)
- M4T**: Status: request submitted, Result: [M4T Website](#)
- HHpred**: Status: request submitted, Results will be sent by email.
- Phyre2**: Status: request submitted, Result: [Phyre2 Website](#)

Slika 10: Interaktivno modeliranje z različnimi strežniki preko PMP

smo izbrali vse strežnike: HHpred, I-TASSER, ModWeb, M4T, SWISS-MODEL ter Phyre<sup>2</sup>. Za oceno kvalitete modela pa ModEVAL, ModFOLD3 in QMEAN. Vhodne nastavitev so bile pri vseh strežnikih privzete s strani portala. Pri strežniku I-TASSER je bila potrebna registracija, zato smo modeliranje s tem strežnikom izpustili.

Drugič smo za modeliranje zagnali le Phyre<sup>2</sup>. Način modeliranja (»Intensive«) je bil privzet. S programom Phyre<sup>2</sup> smo zmodelirali dva modela, ki smo ju poimenovali TLR15\_0 in TLR15\_1.

### 3.5 NEPOSREDNO MODELIRANJE S STREŽNIKOM PHYRE<sup>2</sup>

Na domači strani strežnika so polja za vnos elektronskega naslova, kamor strežnik posreduje rezultate, polje za vnos opisa dela ter polje za vnos aminokislinske sekvence. Na koncu je še možnost izbere načina modeliranja (»Normal« ali »Intensive«) (slika 11).

V prejšnjem koraku smo za modeliranje TLR15 preko PMP posredno uporabili ta strežnik. Za nestrokovnjake na področju modeliranja je precej enostaven za uporabo. Oba načina Phyre<sup>2</sup> smo uporabili še posredno za modeliranje: v načinu »Normal« smo model poimenovali TLR15\_2, v načinu »Intensive« pa TLR15\_3, TLR15\_4, TLR15\_5, TLR15\_6. Modele posrednega in neposrednega modeliranja smo primerjali med sabo.



Slika 11: Domača stran strežnika Phyre<sup>2</sup>

### 3.6 PRIMERJAVA MODELOV ZMODELIRANIH Z RAZLIČNIMI STREŽNIKI TER ANALIZA MODELA MOLEKULE TLR15

S programom Pymol (slika 12) smo grafično primerjali modele, ki so nam jih izračunali različni strežniki. Primerjali smo modele izračunane posredno z interaktivnim modeliranjem s pomočjo PMP, torej modele napovedane s strežniki HHpred, SWISS-MODEL, ModWeb in Phyre<sup>2</sup> ter neposredno s strežnikom Phyre<sup>2</sup>.

S programom smo odprli vsako molekulo posebej in glede na iste/različne koordinate, s katerimi so določeni atomi molekul, smo iz slike modela lahko razbrali, kje so modeli isti, podobni oziroma različni.

### 3.6.1 Primerjava modelov zmodeliranih s strežnikom Phyre<sup>2</sup> (posredno s PMP in neposredno)

Najprej smo primerjali modela TLR15\_0 in TLR15\_1, ki sta bila izračunana posredno s PMP. Pogledali smo na katerih mestih sta modela enaka/podobna oziroma različna. Nada- lje smo na isti način primerjali še modele TLR15\_0 ter TLR15\_3 in TLR15\_4, ki sta bila izračunana neposredno s strežnikom Phyre<sup>2</sup>.

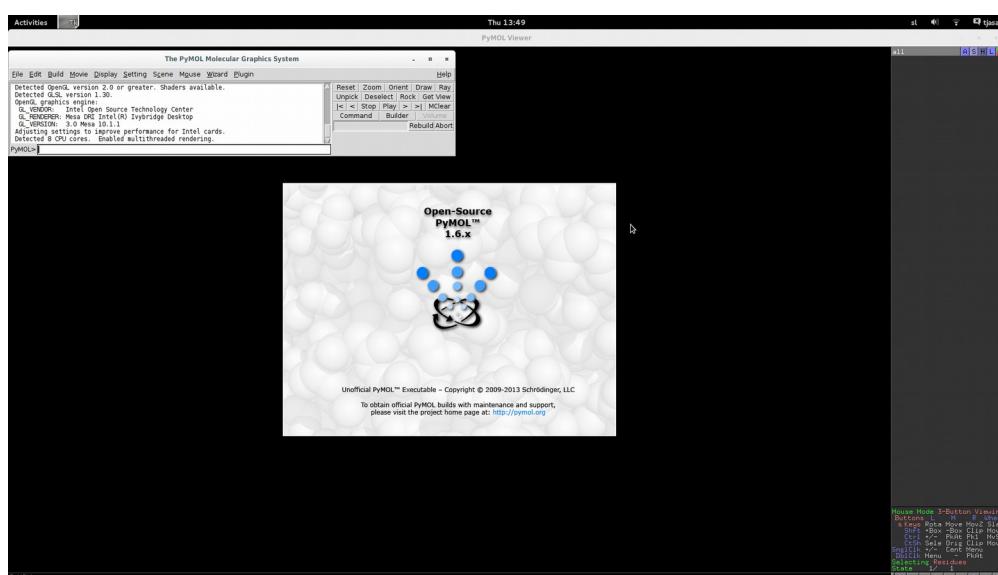
### 3.6.2 Primerjava modelov zmodeliranih interaktivno s PMP (HHpred, SWISS-MODEL, ModWeb in Phyre<sup>2</sup>)

Med sabo smo primerjali tudi modele, interaktivno zmodelirane z različnimi strežniki. Primerjali smo modele celotnih molekul, levcinskih ponovitev ter domene TIR. Model izračunan s strežnikom HHpred smo s programom Pymol ločili na domeno levcinskih ponovitev in domeno TIR. SWISS MODEL strežnik nam je sam po sebi izračunal model celotne molekule, model domene levcinskih ponovitev ter model domene TIR. Strežnik ModWeb nam ni izračunal modela celotne molekule ter levcinskih ponovitev, zato smo za primerjavo uporabili le model domene TIR.

Primerjave so torej bile naslednje: modeli TLR15 izračunani s strežniki HHpred, SWISS-MODEL ter Phyre<sup>2</sup> (model TLR15\_1), modeli levcinskih ponovitev izračunani s strežniki HHpred, SWISS MODEL, Phyre<sup>2</sup> (ektodomena modela TLR15\_6) ter modeli domene TIR izračunani s strežniki HHpred, SWISS-MODEL, ModWeb ter Phyre<sup>2</sup> (TIR od TLR\_6).

### 3.6.3 Analiza napovedanih domen modela molekule TLR15

Pymol je grafični vmesnik za prikazovanje in urejanje modelov molekul. Omogoča nam označevanje, izrezovanj, kopiranje, itd. delov molekul.



**Slika 12:** Pymol - program za grafični prikaz in urejanje molekul

S programom Pymol (slika 12) smo za boljšo predstavo in lažjo analizo vezavnih mest, na modelu TLR15\_6 (zmodeliran s strežnikom Phyre<sup>2</sup>) z različnimi barvami označili posamezne levcinske ponovitve.

### 3.7 NAPOVED ŽEPOV, VOTLIN/LUKENJ TER VEZAVNIH MEST ZA LIGANDE

Za napoved in iskanje žepov, votlin/lukenj ter vezavnih mest smo uporabili zadnji model izračunan s programom Phyre<sup>2</sup> (TLR15\_6).

#### 3.7.1 Napoved vezavnega mesta za ligande z meta-strežnikom COACH

Na spletni strani strežnika (slika 13) smo najprej naložili datoteko .pdb s strukturo modela TLR15 zmodeliranega s strežnikom Phyre<sup>2</sup> (TLR15\_6). Pozneje smo napoved zagnali še enkrat, vendar z datoteko .txt s formatom FASTA sekvence proteina TLR15 (C4PBP0). Vnesli smo elektronski naslov za posredovanje rezultatov ter zagnali napoved. Rezultate napovedi smo primerjali med seboj. Za boljšo preglednost smo kompleksom ligand-protein odstranili regijo TIR.

The screenshot shows the COACH web interface. At the top, there is a logo featuring a 3D molecular structure on the left and the text "COACH" in large bold letters, with "for" and "Protein-ligand binding site prediction" below it. To the right of the text are icons representing a molecule, a green sphere, and several small colored spheres. Below the header, there is a text input field with placeholder text: "Please copy and paste your data below (either structure in [PDB format](#) or sequence in [FASTA format](#) is acceptable)". Underneath this, there are two links: "Click for an example PDB input" and "Click for an example FASTA input". A large, empty text area follows, which is where the user would paste their data. Below this area, there is a section for file upload with the text "Or upload the stucture/sequence file:" and a "Choose File" button. The file selection dropdown shows "No file chosen". Further down, there are fields for "Email: (Optional, where results will be sent to)" and "ID: (Optional, your given name to this protein)". At the bottom of the form are two buttons: "Run COACH" and "Clear form".

Slika 13: Domača stran strežnika COACH

## 4 REZULTATI

### 4.1 OBSTOJEČA ZAPOREDJA MOLEKULE TLR15 PRIDOBELJENA S POMOČJO PODATKOVNE BAZE UNIPROT

S pomočjo iskalnika v podatkovni zbirki UniProt smo pridobili vse obstoječe sekvence molekule TLR15. Dobili smo 63 zadetkov, ki ustrezajo imenu proteina TLR15 in pripadajo 13 različnim organizmom. Oblikovali smo preglednico, kjer smo število in dolžino razpoložljivih sekvenčnih zaporedij po imenih organizmov, ki jim pripadajo (preglednica 2).

Genomska sekvenca je bila najprej določena kokoši (*Gallus gallus*) leta 2004. Izbrali smo eno sekvenco vsakega organizma, ga izvozili v formatu FASTA ter shranili kot datoteko .txt. Te smo nato uporabili za poravnavo sekvenčnih zaporedij s programom BLAST. Naredili smo tudi datoteko .txt, ki vsebuje format FASTA vseh 13 zaporedij in to datoteko uporabili za poravnavo s programom Clustal Omega. Vsako od teh zaporedij je potencialna tarča za homologno modeliranje molekule TLR 15.

**Preglednica 2:** Razpoložljiva zaporedja TLR15 v proteinski podatkovni zbirki UniProt

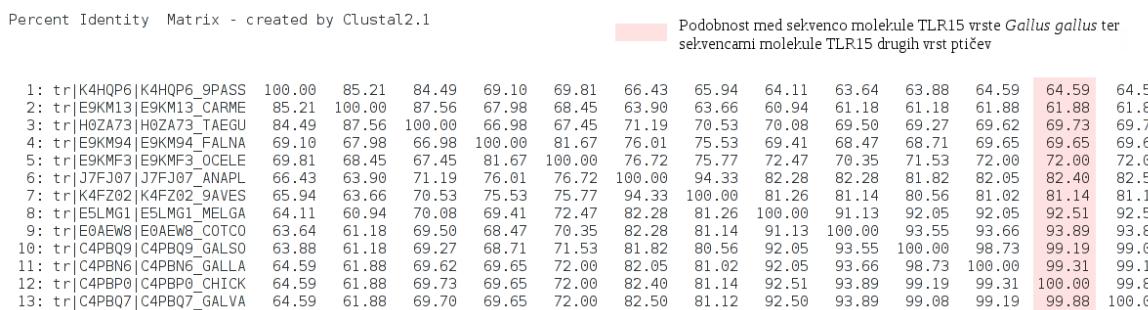
Ime organizma	Ime vnosa	Število zaporedij	Dolžina AK zaporedja
<i>Anas platyrhynchos</i>	J7FJ07_ANAPL	1	864
<i>Anser anser</i>	K4FZ02_9AVES	1	865
<i>Carpodacus mexicanus</i>	E9KM13_CARME	16	433
<i>Coturnix coturnix</i>	E0AEW8_COTCO	2	868
<i>Falco naumanni</i>	E9KM94_FALNA	13	431
<i>Gallus gallus</i>	C4PBP0_CHICK	16	868
<i>Gallus lafayetii</i>	C4PBN6_GALLA	1	868
<i>Gallus sonneratii</i>	C4PBQ9_GALSO	4	868
<i>Gallus varius</i>	C4PBQ7	1	868
<i>Meleagris gallopavo</i>	E5LMG1_MELGA	1	868
<i>Oceanodroma leucorhoa</i>	E9KMF3_OCELE	1	431
<i>Petroica australis rakiura</i>	K4HQP6_9PASS	3	426
<i>Taeniopygia guttata</i>	H0ZA73_TAEGU	1	867

Z orodjem Clustal Omega in BLAST smo poravnali razpoložljiva zaporedja TLR15, ki pripadajo različnim redom, družinam in vrstam.

### 4.2 PORAVNAVA OBSTOJEČIH SEKVENC MOLEKULE TLR15

#### 4.2.1 Poravnavo več sekvenčnih zaporedij s programom Clustal Omega

S tem programom smo poravnali vseh 13 zaporedij molekule TLR15, kjer smo uporabili datoteko .txt, ki je vsebovala vseh 13 zaporedij. Slika 14 prikazuje podobnost v % med sekvenčnimi molekule TLR15 med različnimi organizmi. Na sliki lahko vidimo, da so si med sabo najbolj podobne (nad 95 %) molekule TLR organizmov rodu *Gallus* (*gallus*, *varius*, *lafayetii* ter *sonneratii*).



Slika 14: Rezultati poravnave več zaporedij

#### 4.2.2 Poravnave dveh sekvenc s programom BLAST

S tem programom smo poravnali zaporedje TLR15 domačega petelina (*Gallus gallus*) z zaporedji TLR15 ostalih ptičev. Naredili smo torej 13 poravnav ter dobili rezultate, prikazane v preglednici 3.

Preglednica 3: Rezultati poravnav s programom BLAST

Vhodna sekvenca (angl. Query)	Zaporedje iz baze (angl. Subject)	Podobnost (%)	Rezultat (angl. Score)	E-vrednost
<i>Gallus gallus</i>	<i>Anas platyrhynchos</i>	81	1417	0,0
<i>Gallus gallus</i>	<i>Anser anser</i>	80	1402	0,0
<i>Gallus gallus</i>	<i>Carpodacus mexicanus</i>	61	489	8e-169
<i>Gallus gallus</i>	<i>Coturnix coturnix</i>	94	1657	0,0
<i>Gallus gallus</i>	<i>Falco neumanni</i>	69	562	0,0
<i>Gallus gallus</i>	<i>Gallus lafayetii</i>	99	1796	0,0
<i>Gallus gallus</i>	<i>Gallus sonneratii</i>	99	1765	0,0
<i>Gallus gallus</i>	<i>Meleagris gallopavo</i>	93	1660	0,0
<i>Gallus gallus</i>	<i>Gallus varius</i>	99	1779	0,0
<i>Gallus gallus</i>	<i>Oceanodroma leucorhoa</i>	71	687	0,0
<i>Gallus gallus</i>	<i>Petroica australis rakiura</i>	63	503	2e-174
<i>Gallus gallus</i>	<i>Taeniopygia guttata</i>	68	1186	0,0

\*Vhodna sekvenca – opisno zaporedje s katerim iščemo po bazi

Podobne rezultate kot v prejšnji poravnavi smo dobili tudi pri primerjanju molekule TLR organizma *Gallus gallus* z ostalimi organizmi. Največja podobnost (99 %) je med molekulami organizmov rodu *Gallus* (*lafayetii*, *sonneratii* in *varius*). Več kot 90 % podobnost ima molekula TLR15 organizma *Gallus gallus* še z organizmoma *Coturnix coturnix* ter *Meleagris gallopavo*.

### 4.3 ANALIZA SEKVENCE MOLEKULE TLR15 S PODATKOVNO BAZO TOLLML

#### 4.3.1 Primerjava izbrane sekvence (C4PBP0) z vsemi sekvencami v podatkovni bazi

Primerjava naše sekvence s sekvencami v podatkovni bazi TollML je dala naslednje rezultate: sekvenca se z E-vrednostjo 0 in rezultatom 3617 ujema z vnosni, ki so sekvence molekul TLR15. Rezultati prvih desetih poravnav so prikazani na sliki 15.

**Report :**

- Sequence name : TLR15
- Sequence length : 868
- Method : blastp
- Database : whole TollML DB
- Scoring matrix : pam70

**Result :**

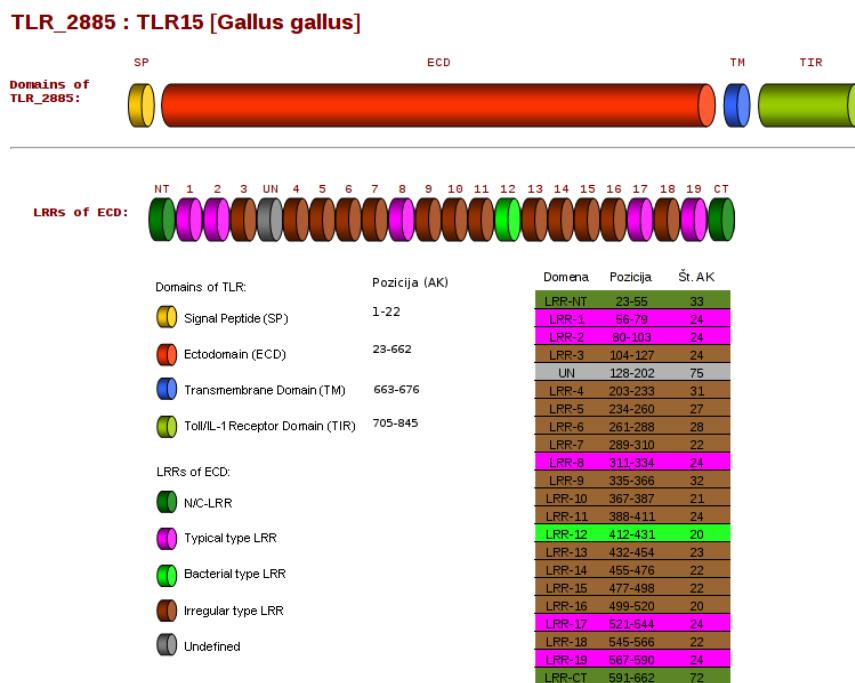
Multiple Alignment					
Nr.	TollML ID	E-value	Score	TollML Label	Links
0 <input checked="" type="checkbox"/>	Query Sequence name : TLR15				
1 <input type="checkbox"/>	TLR_2885	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
2 <input type="checkbox"/>	TLR_2886	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
3 <input type="checkbox"/>	TLR_2887	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
4 <input type="checkbox"/>	TLR_2888	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
5 <input type="checkbox"/>	TLR_2889	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
6 <input type="checkbox"/>	TLR_2897	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
7 <input type="checkbox"/>	TLR_2912	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
8 <input type="checkbox"/>	TLR_2892	0.00000	3617	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
9 <input type="checkbox"/>	TLR_2878	0.00000	3613	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>
10 <input type="checkbox"/>	TLR_1009	0.00000	3611	TLR15;TLR	<a href="#">Sequence</a> , <a href="#">Domains</a> , <a href="#">LRRs</a>

Slika 15: Rezultati primerjave sekvence TLR15 s sekvencami v podatkovni bazi TollML

#### 4.3.2 Domene molekule TLR15 določene s programom LRRFinder

Naša molekula je tako kot vse ostale molekule TLR sestavljena iz signalnega peptida, ekto-domene oziroma domena levcinskih ponovitev, transmembranske domene ter domene TIR. Signalni peptid, se nahaja na poziciji 1-22 AK, ektdomena na pozicij 23-662 AK, transmembranska domena na poziciji 663-676 ter domena TIR na poziciji 705-845. Ektodomena vsebuje 21 LRR domen ter 75 AK dolgo nedefinirano zaporedje.

Začetek in konec ektdomena predstavlja N-terminalna levcinska ponovitev (LRR-NT) ter C-terminalna levcinska ponovitev (LRR-CT), LRR-1,2,8,17 in 19 predstavljajo tipični tip, LRR-12 predstavlja bakterijski tip, LRR-3,4,5,6,7,9,10,11,12,13,14,15,16,18 pa predstavljajo neobičajni tip LRR-jev. Domene so shematsko prikazane na sliki 16.



Slika 16: Domene in pozicije domen molekule TLR15

**Preglednica 4:** Pozicije in sekvene domen molekule TLR15

Domena	Pozicija	Sekvenca (LxxLxLxxNxL ...)	
SP	1-22	MRILIGSLYFYFISFLFSKVNG	
LRR-NT	23-55	FLTQRTSPVSSFPFYNSYLNSSVSQAQAPKT	
LRR1	56-79	ARALNFSYNNAI	EKITKRDSEGPHV
LRR2	80-103	LEVLDLSHNHI	KDIEPGAFENLLS
LRR3	104-127	LVSVDLSFNDK	NLLVSGLAPHLKL
UN	128-202	IPTSGASGPSQIYMYFQKSAEAALEPSAPAELLPHLEDPPNPNGNVNPFRQRTE-ENKTSPPAATLRPDLCGAPI	
LRR4	203-233	NGLLDLSRTKL	SNEELTAKLDADLCQAQLGT
LRR5	234-260	VLEFNISHSDL	EMDLLSLFILFLPMKD
LRR6	261-288	IQSVDASYNRI	TINNIDVEAICHFPFSN
LRR7	289-310	FSFLNISNNPI	NSLETVCLPAS
LRR8	311-334	ITVIDLSFTNI	STIPANFAKKLSK
LRR9	335-366	LERMYVQGNQL	IYTVRPENPSATPRPPPGTVQ
LRR10	367-387	ISAISLVRNQA	GTPIESLPES
LRR11	388-411	VKHLKVSNCASI	VELPEWFANRMQE
LRR12	412-431	LLFLDLSSNRI	SMLPDLPIS
LRR13	432-454	LQQLDISNSDI	KIIPPRFKSLSN
LRR14	455-476	LTVFNIQNNKL	TEMHPEYFPST
LRR15	477-498	LTTCDISKNKL	KVLSLTAKLEN
LRR16	499-520	LESLNVSGNLI	TRLEPACQLPS
LRR17	521-544	LTNLDDSSHNL	SELPDHGQSLLM
LRR18	545-566	LKHFNLSGNKI	SFLQRGSLPAS
LRR19	567-590	LEELDISDNAI	TTIVQDTFGQLTS
LRR-CT	591-662	LSVLTVQGKHFFCNCDLYWFVNIYIRNPHLQINGKDDLRCSFPPDRRGSLVKS-SNLTLHCSLGQMAITAC	
TM	663-676	MAILVVVLVTGLCWR	
CD/TIR	677-868/ 705-845	RFDGLWYVRMGWYWCMAKRRQYKKRPFDAFISYSEHDADWTKE-HLLKKLETDGFKICYHERDFKPGHPVGNIFYCIENSHKVLFLSPSFVN-SCWCQYELYFAEHRVLDENQDSLIMVVLEDLPPDSVPQKFSKLRKLLKRK-TYLKW SPEEHKQKIFWHQLAAVLKTTNEPLVRAENGPNEDVIEME	

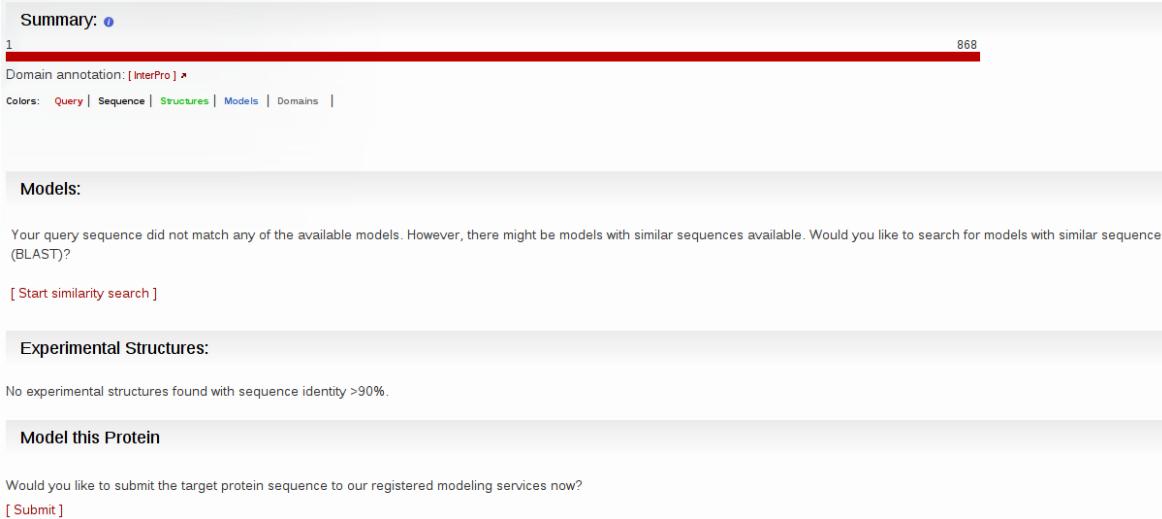
Preglednica 4 poleg pozicij prikazuje še sekvene posameznih domen. SP predstavlja signalni peptid molekule. Pri LRR1 – LRR19, 3. in 4. stolpec v preglednici vsebujeta sekveno ene levcinske ponovitve – v 3. stolpcu so sekvene regije HCS levcinskih ponovitev, v 4. pa regije VS. LRR-NT in LRR-CT ne vsebujeta HCS in VS. Vrstica CD vsebuje celotno sekveno citoplazemske domene, znotraj katere se nahaja sekvenca domene TIR (označena z rumeno).

## 4.4 EKSPERIMENTALNE STRUKTURE IN TEORETIČNI MODELI MOLEKULE TLR15 IN MOLEKULE S PODOBNIMI SEKVENCAMI

### 4.4.1 Eksperimentalne strukture in teoretični modeli na voljo na spletu

Z orodjem Protein Model Portal smo poskušali poiskati že obstoječe modele ter eksperimentalne strukture molekule TLR15, vendar naši sekvenci ni ustrezal noben model, niti eksperimentalna struktura, ki bi imela podobnost z našo molekuljo večjo od 90 % (slika 17).

#### Query Result:



Slika 17: Rezultati iskanja eksperimentalnih struktur in teoretičnih modelov z uporabo sekvence TLR15

### 4.4.2 Iskanje modelov s podobnimi sekvencami

Nadalje nam je strežnik ponudil opcijo, da s pomočjo algoritma BLAST poiščemo modele s podobnimi sekvencami (slika 17). BLAST je našel strukturno informacijo 12 modelov, ki ustrezajo proteinu TLR15 (slika 18).

Summary:			
Structural information is available for the following Proteins:			
Link to Models	Protein Description	Exp. Structures	Models
C4PBN6	C4PBN6_GALLA Unreviewed; 868 AA. TLR15; Gallus lafayettei (Ceylon junglefowl).	0	2
C4PBP2	C4PBP2_CHICK Unreviewed; 868 AA. TLR15; Toll-like receptor 15; Gallus gallus (Chicken).	0	1
C4PBQ9	C4PBQ9_GALSO Unreviewed; 868 AA. TLR15; Gallus sonneratii (Gray junglefowl).	0	2
C4PBR0	C4PBR0_GALSO Unreviewed; 868 AA. TLR15; Gallus sonneratii (Gray junglefowl).	0	2
C4PBS4	C4PBS4_CHICK Unreviewed; 868 AA. TLR15; Gallus gallus (Chicken).	0	1
E5LMG1	E5LMG1_MELGA Unreviewed; 868 AA. Toll-like receptor 15; Meleagris gallopavo (Common turkey).	0	2
Q2XQ10	Q2XQ10_CHICK Unreviewed; 868 AA. Toll-like receptor 15; Gallus gallus (Chicken).	0	2

Slika 18: Strukturne informacije modelov s podobnimi sekvencami, ki so na voljo

Preglednica 5 prikazuje informacije o modeliranju vseh 12 modelov: vrsto ptiča, katerega sekvenca molekule je bila uporabljena za modeliranje, vnos UniProt ter dolžino teh sekvinc; šablonu uporabljenou za modeliranje; sekvenčno podobnost med vneseno sekvenco

in sekvenco šablone; katera domena oziroma del molekule je zmodeliran; dolžino zmodeliranega dela molekule ter datum modeliranja.

**Preglednica 5:** Modeli napovedani iz zaporedja TLR 15, ki so na voljo na spletu

Vnesene sekvence	Šablona	Podobnost (%)	Domena	Dolžina modela (AK)	Datum nastanka
<i>Gallus lafayettei</i> (C4PBN6), 868 AK	1FYW_A	47	TIR	143 (704-847)	Maj 2010
	2JS7_A	23		149 (698-847)	Junij 2010
<i>Gallus gallus</i> (C4PBP2), 868 AK	1VL5_A	49	del TIR	38 (702-740)	April 2011
<i>Gallus Sonneratti</i> (C4PBQ9), 868 AK	1FYW_A	47	TIR	143 (704-847)	Maj 2010
	2JS7_A	23		149 (698-847)	Junij 2010
<i>Gallus Sonneratti</i> (C4PBR0), 868 AK	1FYW_A	47	TIR	143 (704-847)	Maj 2010
	2JS7_A	23		149 (698-847)	Junij 2010
<i>Gallus gallus</i> (C4PBS4), 868 AK	1VL5_A	49	del TIR regije	38 (702-740)	September 2011
<i>Melleagris gallopavo</i> (E5LMG1), 868 AK	1ZIW_A	25	LRR-ji	572 (40-612)	Oktober 2011
	1ZIW_A	10	LRR-ji + TIR	292 (549-841)	Oktober 2011
<i>Gallus gallus</i> (Q2XQ10), 868 AK	1FYW_A	47	TIR	143 (704-847)	Maj 2010
	2JS7_A	23		149 (698-847)	Junij 2010

Sekvence za izdelavo modelov (angl. *query*) so bili različni vnosi (iz podatkovne zbirke UniProt) molekule TLR15, enakih ali različnih vrst ptičev. 10 modelov vsebuje struktorno informacijo domene TIR (8 od teh strukturo celotne domene, 2 pa le del domene), en model vsebuje struktorno informacijo domene LRR-jev ter en model struktorno informacijo končnega dela domene LRR-jev (nekaj levcinskih ponovitev) z domeno TIR. Oba modela LRR-jev sta bila napovedana za molekulo TLR15 purana (*Maleagris gallopavo*), vsi ostali modeli (domene TIR) pa za molekulo TLR 15 različnih vrst rodu Gallus (med različnimi vrstami je podobnost sekvence >95 %). Za modeliranje celotne domene TIR, sta bili uporabljeni dve šabloni, in sicer molekula TLR2 (PDB ID: 1FYW) ter molekula MyD88 (PDB ID: 2JS7). Za modeliranje 38 AK dolgega dela domene TIR pa je bila uporabljena ena šablona in sicer domnevna metiltransferaza bakterije *Bacillus Halodurans* (PDB ID: 1VL5). Za modeliranje levcinskih ponovitev je bila uporabljena šablona molekula TLR3 (PDB ID: 1ZIW).

Pravilnost teh modelov je odvisna od podobnosti med tarčo in šablono. Glede na to, da so si sekvence TLR15 vseh ptičev rodu *Gallus* ter *Maleagris* (ki pripadajo isti družini) zelo podobne (nad 90 %), imajo le-te enako podobnost z določenimi šablonami. Naša tarčna molekula ima v vseh primerih s šablono 1FYW (TIR domena TLR2) 47 % podobnost, s šablono 2JS7 (MyD88) 23 % podobnost, s šablono 1VL5 (BH2331) 49 % podobnost ter s šablono 1ZIW v primeru enega modela 25 % podobnost in v primeru drugega modela 10 % podobnost.

Strežnik PMP nam je poleg iskanja modelov s podobnimi sekvencami ponudil tudi možnost interaktivnega modeliranja molekule z različnimi strežniki.

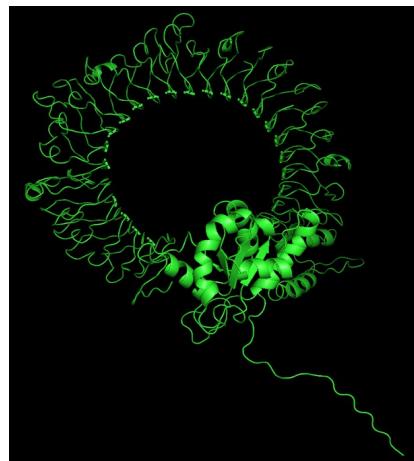
## 4.5 INTERAKTIVNO (POSREDNO) MODELIRANJE S POMOČJO PMP

### 4.5.1 Model izračunan s strežnikom HHpred

Končni model nam je strežnik po elektronski pošti posredoval kot izvorno datoteko v TS kodi formata (datoteka, ki je v PDB formatu vsebuje podatke o 3D atomskih koordinatah). Shranili smo jo s končnico .pdb, da smo jo lahko odprli in pregledali s programom Pymol. V izvorni datoteki so bili le podatki o metodi ter šablonah na podlagi katerih je bil izračunan model. Šablone so naštete v preglednici 6, model pa prikazan na sliki 19.

**Preglednica 6:** Informacije o šablonah, uporabljenih za modeliranje s strežnikom HHpred

PDB ID	Molekula	Del molekule
3J0A_A	TLR5	Celotna molekula
1ZIW_A	TLR3	Ektodomena
3ZYI_A	NetringG2	Levcinske ponovitve
2J67_A	TLR10	Domena TIR
2FT3_A	Biglycan dimer	Levcinske ponovitve
1FYX_A	TLR2	Domena TIR
2Z66_A	TLR4	Levcinske ponovitve
1FYV_A	TLR1	Domena TIR
4G8A_B	TLR4	Levcinske ponovitve



**Slika 19:** Model TLR15 izračunan s strežnikom HHpred

### 4.5.2 Modeli izračunani s strežnikom SWISS-MODEL

V tem primeru nam je strežnik po elektronski pošti posredoval povezavo do rezultatov na spletni strani in ne končni model kot v prejšnjem primeru. Strežnik nam je izračunal 9 modelov, vsakega s pomočjo ene šablone. Modeli zajemajo različno število in delež AK sekvenci molekule TLR15. Slika 20 prikazuje kateri deli sekvence molekule so bili zmodelirani (modra), glede na celotno sekenco, ki smo jo vnesli (zelena).



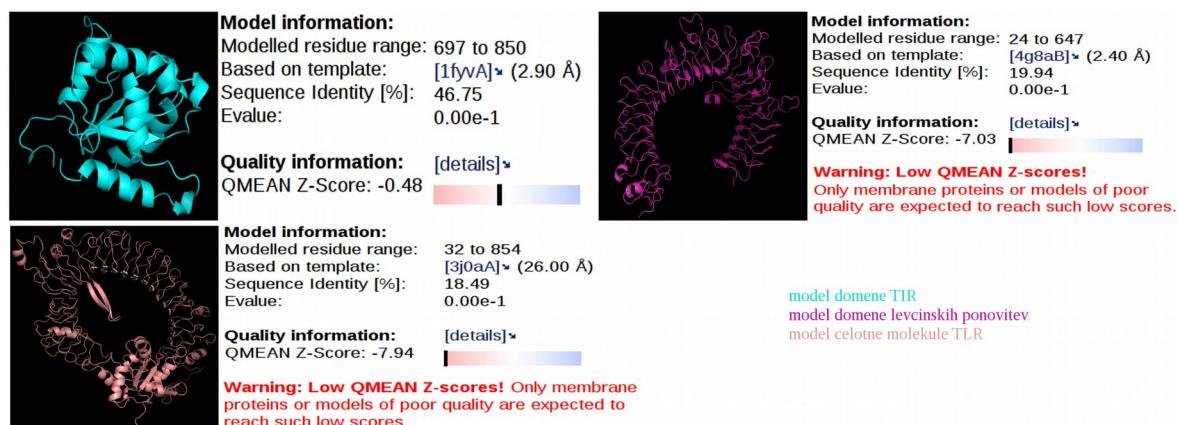
Slika 20: Modeli izračunani s strežnikom SWISS-MODEL

Za vsak posamezni model so podane tudi dodatne informacije, ki so navedene v preglednici 7: kateri del molekule je zmodeliran, dolžina zmodelirane sekvence, PDB ID šablone, sekvenčna podobnost med tarčo in šablono (vrednosti so bile podane na dve decimalni mestni natančno in smo jih zaokrožili na celo število) ter informacijo o kvaliteti modela. 6 modelov vsebuje strukturno informacijo o kratkih segmentih domene levcinskih ponovitev, od ostalih 3 modelov pa en predstavlja celotno molekulo TLR, en domeno TIR in en celotno domeno levcinski ponovitev. V preglednici so z rumeno označeni modeli, ki so prikazani na sliki 21.

Preglednica 7: Modeli izračunani s strežnikom SWISS-MODEL

Domena	Dolžina modela (AK)	Šablona	Podobnost (%)	Q-MEAN
LRR-ji	104 (26-130)	3ZYI_A	33	-2,23
LRR-ji	37 (309-346)	3W3G_A	24	-0,71
LRR-ji	217 (385-602)	1JL5_A	29	-3,65
TIR	153 (697-850)	1FYV_A	47	-0,48
LRR-ji	195 (456-651)	2V9T_B	26	-2,15
LRR-ji	316 (288-604)	3G06_A	23	-4,56
LRR-ji	237 (412-649)	2O6Q_A	22	-3,04
LRR-ji	623 (24-647)	4G8A_B	20	-7,03
TLR	822 (32-854)	3J0A_A	19	-7,94

Slika 21 prikazuje modele domene TIR (modra) ter domene levcinskih ponovitev (vijolična) in model celotne molekule (roza).



Slika 21: Modeli ter dodatne informacije izračunani s strežnikom SWISS-MODEL

Za šablono za modeliranje domene TIR je strežnik izbral kristalno strukturo domene TIR človeške molekule TLR2 (PDB ID: 1FYV), s katero ima sekvenca naše molekule 47 % podobnost, model pa srednje dober Q-MEAN (-0,48). Za šablono za modeliranje domene levčinskih ponovitev je strežnik izbral kristalno strukturo človeške molekule TLR4 (PDB ID: 4G8A) s katero ima sekvenca naše molekule 20 % podobnost, model pa nizek Q-MEAN (-7,03). Za šablono za modeliranje celotne molekule TLR15 je strežnik izbral model človeške molekule TLR5 (PDB ID: 3J0A), s katero ima sekvenca naše molekule 18 % podobnost, model pa nizek Q-MEAN (-7,94).

#### 4.5.3 Model izračunan s strežnikom ModWeb

Strežnik ModWeb je izračunal 4 modele ter nam posredoval informacije po elektronski pošti. Za modeliranje je pri vsakem modelu izbral eno šablono. Nobeden od modelov ni celotna molekula TLR. En od modelov je domena TIR in je prikazan na sliki 22. V preglednici 8 so navedene dodatne informacije o modelih: kateri deli molekule so bili zmodelirani, dolžina zmodelirane sekvence, PDB ID šablone, sekvenčna podobnost med tarčo in šablono ter informacije o kvaliteti modela.

**Preglednica 8:** Modeli ter dodatne informacije o modelih izračunanih s strežnikom ModWeb

Domena	Dolžina modela (AK)	Šablona	Podobnost (%)	Kvaliteta modela
LRR-ji	148 (431-580)	1d0bA	31	1
LRR-ji	334 (262-596)	4kt1A	27	1
LRR-ji	63 (59-112)	2r9uA	41	0,95
TIR	142 (704-846 AK)(TIR domena)	1fyvA	50	1



**Slika 22:** Model TIR domene izračuna s strežnikom ModWeb

#### 4.5.4 Model izračunana s strežnikom Phyre<sup>2</sup>

PMP avtomatsko izbere modeliranje v načinu »Intensive«. Strežnik nam po elektronski pošti posreduje izračunan model celotne molekule (v formatu .pdb) ter povezavo do rezultatov na strežniku, kjer so poleg modela prikazane še dodatne informacije o modelu ter modeliranju.

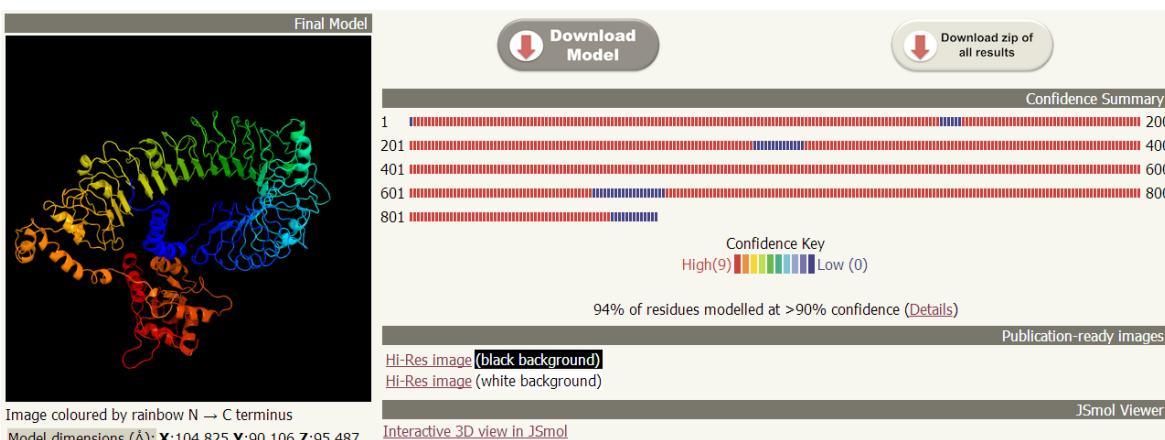
Rezultati na strani strežnika so prikazani v več razdelkih:

- povzetek (slika modela in shematski prikaz zaupanja)
- sekvenčna analiza (večsekvenčna poravnava Psi-Blast)

- napoved sekundarne strukture in neurejenosti
- analiza domen
- podrobna informacija o modelih ter njihovih šablonah
- informacije o modeliranju *ab initio*; prikaz pokritost sekvence z izbranimi šablonami ter celokupna zaupljivost ujemanja sekvence z izbranimi šablonama

Razen samega modela molekule, se rezultati (dodatne informacije o modelu ter modeliranju) za oba modela TLR15\_0 in TLR15\_1, zelo malo razlikujejo, zato so v nalogi prikazane le slike z rezultati za model TLR15\_1, razlike med modeloma pa so dodatno navedene. Primerjava modelov je prikazana v razdelku 4.7.2.

Slika 23 poleg skice modela prikazuje tudi informacijo o zaupanju modela – 94 % AK ostankov je zmodeliranih z več kot 90 % zaupanjem, ostalih 6 % pa z zelo nizkim zaupanjem (zaupljivost modela TLR15\_0 je identična). Zaupljivost je definirana s ključem (angl. *confidence key*), ki ima barvno lestvico, kjer rdeča pomeni največje in modra najmanjše zaupanje.



Slika 23: Model TLR15\_1 in zaupljivost

V razdelku, kjer sta napovedani sekundarna struktura ter neurejenost, je prikazana tudi zaupljivost napovedane sekundarne strukture ter neurejenosti. Sekundarno strukturo sestavlja jo 3 oblike stanja:  $\alpha$ -vijačnica,  $\beta$ -list ter zanka. Oba modela TLR15\_0 in TLR15\_1 sta sestavljena iz 20 %  $\alpha$ -vijačnic, 13 %  $\beta$ -listov ter 67 % zank. Neurejenih regij v proteinu je 32 %.

Razdelek o analizi domen prikazuje pokritost in zaupanje tarčne sekvence s sekvencami različnih šablon. Za 20 šablon, pri katerih sta pokritost in zaupanje največja, je strežnik iz-



Slika 24: Analiza domen za model TLR15\_1

računal modele (model celotne molekule ali model domene TIR ali pa model domene LRR-jev), na podlagi katerih je oblikoval končni model. Slika 24 prikazuje PDB ID šablon, pokritost ter zaupljivost za model TLR15\_1.

Pri modelu TLR15\_1 je v 1 primeru s pomočjo šablone zmodelirana cela molekula, v 17 primerih domena LRR-jev ter v 2 primerih samo domena TIR. Pri modelu TLR15\_0 pa v 1 primeru cela molekula (TIR regija in regija z LRR-ji), v 16 primerih domena LRR-jev ter v 3 primerih samo domena TIR. PDB ID-ji šablon so našteti v preglednici 9.

Razdelek s podrobno informacijo o šablonah vsebuje preglednico, ki nam prikaže 3D model zmodeliranega dela molekule, PDB ID šablone, pokritost s šablono (kjer imamo možnost videti tudi poravnavo), zaupljivost, podobnost (v %) s šablono ter informacijo o šabloni (podatki iz zbirke v kateri se šablon nahaja). V preglednici na strani strežnika so modeli prikazani v istem zaporedju kot pri analizi domen (slika 24). Ker je informacij koliko veliko, rezultati niso prikazani.

Slika 25 podrobneje prikazuje pokritost in zaupanje tarčne sekvene s šablonama C3UN9 in C3J0A. Zaupanje s šablono C3UN9 je 98 %, s šablono C3J0A pa 100 %. 54 AK ostankov (1., 146.-152., 295.-309., 650.-671., 856.-868) je bilo zmodeliranih *ab initio* (na novo).



Slika 25: Pokritost in zaupanje z izbranimi šablonama

## 4.6 MODELI IZRAČUNANI NEPOSREDNO S STREŽNIKOM PHYRE<sup>2</sup>

### 4.6.1 Model izračunan v načinu »Normal«

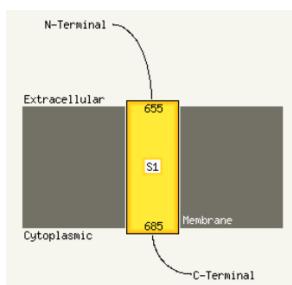
Rezultati so prikazani v razdelkih, enako kot pri modeliranju v načinu »Intensive«, le da pri tem načinu ni razdelka z informacijami o modeliranju *ab initio*, je pa razdelek o napovedi transmembranskih vijačnic.

V primeru modela TLR15\_2 je s pomočjo šablone TLR5 (PDB ID: 3J0A) zmodeliranih 774 ostankov (89 % celotne sekvene) s 100 % zaupanjem. Program je zaznal tudi nekatere dodatne šablonе z zadosti velikim zaupanjem, ki pokrivajo druge regije sekvene. V pri-

meru načina »Intensive« bi tako bilo lahko z uporabo več šablon zmodeliranih 841 ostankov (97 % celotne sekvence) z >90 % zaupanjem.

Rezultati o sekundarni strukturi so enaki kot pri modelih TLR15\_0 in TLR15\_1. Molekulo sestavlja 3 oblike: 20 %  $\alpha$ -vijačnic, 13 %  $\beta$ -listov ter 67 % zank. Neurejenih regij v proteinu je 32 %. Rezultati analize domen, so za izbranih 20 šablon z najboljšo pokritostjo in zaupanjem, enaki kot pri modelu TLR15\_1 (slika 24).

Z neposrednim modeliranjem v načinu »Normal« nam strežnik napove še transmembranske vijačnice. Napovedana topologija proteina je sledeča: N-terminalni del proteina oziraoma prvih 655 AK ostankov (LRR regija) se nahaja izven celice, 30 aminokislinskih ostankov (655-685 AK) se nahaja v membrani ter 183 aminokislinskih ostankov (TIR regija) znotraj celice. Topologija je prikazana na sliki 26.

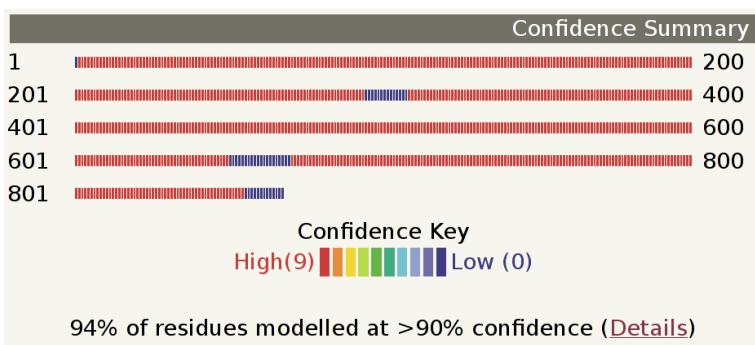


Slika 26: Topologija modela TLR15\_2

#### 4.6.2 Modeli izračunani v načinu »Intensive«

Neposredno s strežnikom Phyre<sup>2</sup> smo zmodelirali še 4 modele: TLR15\_3, TLR15\_4, TLR15\_5, TLR15\_6. Rezultati na spletni strani strežnika so prikazani na enak način, kot že navedeno v razdelku 4.5.4. Pri vseh smo dobili zelo podobne rezultate, kot pri posrednem modeliranju ter modeliranju v načinu »Normal«.

Rezultati v 1. razdelku (skice modelov in zaupanja) so za model TLR15\_4 enaki kot pri modelih TLR15\_0 in TLR15\_1 (slika 23). Pri modelih TLR15\_3, TLR15\_5, TLR15\_6 je kljub procentualno enakemu zaupanju, shematski prikaz zaupanja malo drugačen, ker so *ab initio* zmodelirana 4 različna mesta v modelu (pri modelih TLR15\_0, TLR15\_1 in TLR15\_4 pa 5). Shematski prikaz zaupanja za modele TLR15\_3, TLR15\_5, TLR15\_6 je prikazan na sliki 27.



Slika 27: Zaupanje modela TLR15\_3 (TLR15\_5 in TLR15\_6)

Rezultati analize domen, so za izbranih 20 šablon z najboljšo pokritostjo in zaupanjem, enaki kot pri modelu TLR15\_1 ter TLR15\_2 (slika 24).

Pri vseh 4 modelih sta za prikaz pokritosti in zaupanja tarčne sekvence bili izbrani enaki šabloni, kot pri modelih TLR15\_0 in TLR15\_1 (C3UN9 in C3J0A). Pokritost in zaupanje modela TLR15\_4 s šablonama je enaka kot pri modelih TLR15\_0 in TLR15\_1 (slika 28/levo) in posledično so bile *ab initio* zmodelirane enake regije ter enako število AK ostankov (razdelek 4.5.4). Pri modelih TLR15\_3, TLR15\_5 in TLR15\_6 so bile *ab initio* zmodelirane 4 regije (48 AK ostankov): 1. AK ostanek, 295.-309., 650.-671., 856.-868. AK ostanek (slika 28/desno).

Na sliki 28 so z zeleno označene regije zmodelirane *ab initio*. Z modro je označena regija (dolga 4 AK ostanke), ki pri modelih TLR15\_3, TLR15\_5 in TLR15\_6 niso bili zmodelirani *ab initio* pri modelu TLR15\_4 (ter TLR15\_0 in TLR15\_1) pa.



**Slika 28:** Pokritost in zaupanje modela TLR15\_4 (levo) ter modelov TLR15\_3, TLR15\_5, TLR15\_6 (desno)

V preglednici 9 je prikazan datum modeliranja za posamezni model, način modeliranja ter PDB ID šablon (z rumeno so označene tiste, ki so skupne vsem modelom).

**Preglednica 9:** Informacije o modeliranju s strežnikom Phyre<sup>2</sup>

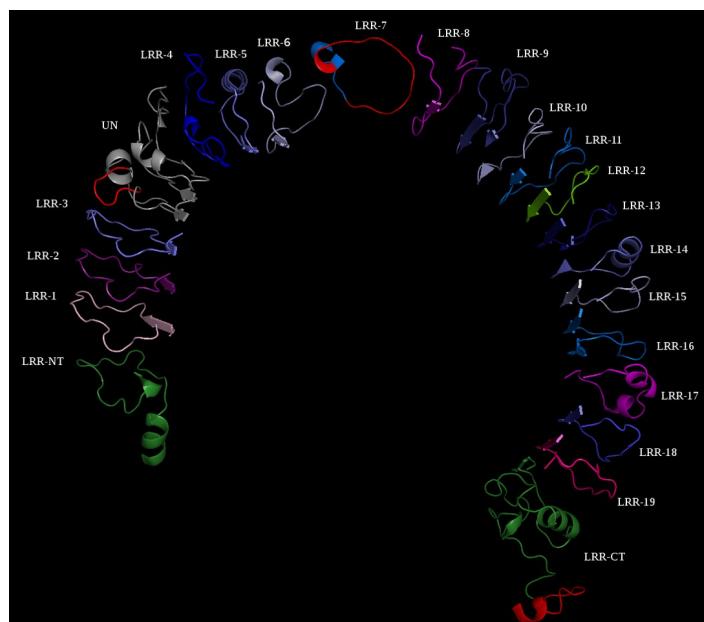
Model	Datum modeliranja	Način	Šablone
TLR15_0	9.11.2013	Intensive*	c3j0aA, c3rg1B, c2z64A, c2a0zA, c3j0aB, c3w3jB, c4g8aB, c3fxiA, c1ziwA, c3a79a, c2z81A, c2z63A, d1fyvA, c3a79B, c2z7xB, c4lsxB, c1t3gB, c3v47B, d1fyxA, c2id5D
TLR15_1	12.12.2013	Intensive*	c3j0aA, c3rg1B, c2z64A, c2a0zA, c3j0aB, c3w3jB, c4g8aB, c3fxiA, c1ziwA, c3a79a, c2z81A, c2z63A, c4mn8A, d1fyvA, c3a79B, c4mn8A, c2z7xB, c4lsxB, c1t3gB, c3v47B
TLR15_2	11.12.2013	Normal	Glej TLR15_1
TLR15_3	12.12.2013	Intensive	Glej TLR15_1
TLR15_4	21.12.2013	Intensive	Glej TLR15_1
TLR15_5	17.1.2014	Intensive	Glej TLR15_1
TLR15_6	20.1.2014	Intensive	Glej TLR15_1

\*Modela sta izračunana posredno preko PMP

## 4.7 ANALIZA DOMENE LEVCINSKIH PONOVITEV MODELA TLR 15 TER PRIMERJAVA MODELOV ZMODELIRANIH Z RAZLIČNIMI STREŽNIKI

### 4.7.1 Grafični prikaz domene levcinskih ponovitev modela molekule TLR15 (TLR15\_6)

Pozicijo in sekvence domen molekule TLR15 smo že določili v razdelku 4.3.2. Na sliki 29 so prikazani LRR-ji v enakem zaporedju kot so razporejeni v molekuli. Obarvani so glede na tip LRR-ja, in sicer s temno zeleno sta obarvana začetni (N-terminalni) LRR in končni (C-terminalni) LRR; z odtenki roza-vijolične barve tipični tip, z odtenki modre neobičajni tip, s svetlo zeleno bakterijski tip sivo pa neidentificirana regija. Z rdečo barvo so označene regije, ki so s programom Phyre<sup>2</sup> zmodelirane *ab initio*.

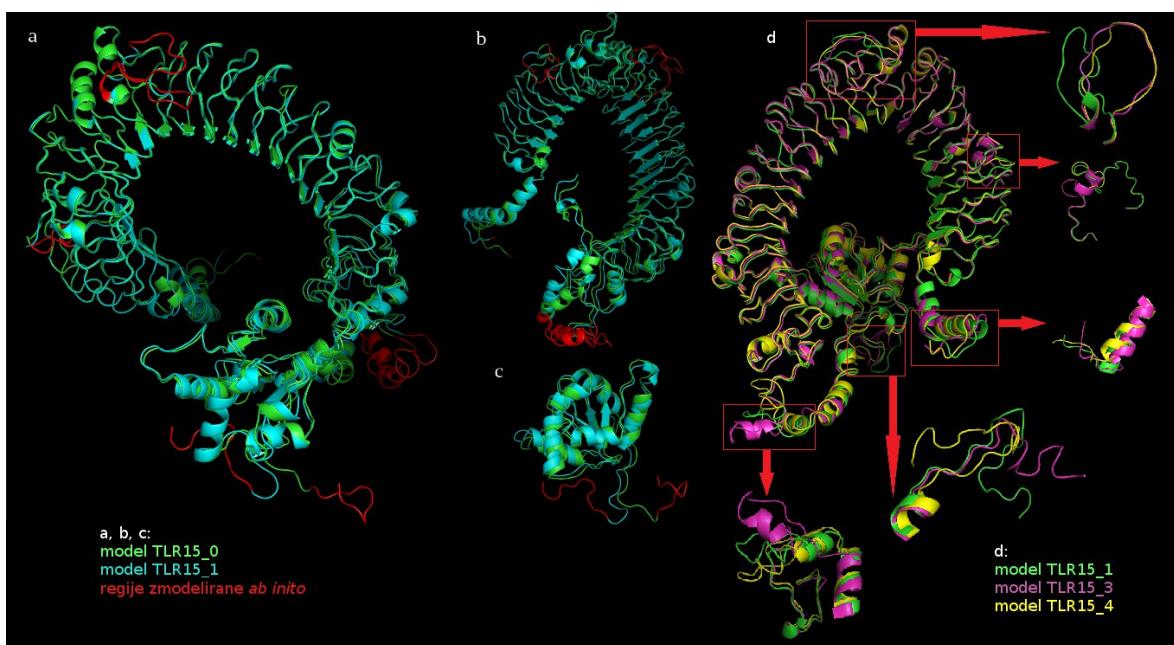


Slika 29: Posamezni LRR-ji domene levcinskih ponovitev

### 4.7.2 Primerjava modelov zmodeliranih s strežnikom Phyre<sup>2</sup>

Slika 30 prikazuje modela TLR15\_0 (zelen) in TLR15\_1 (moder). Razen na 5 mestih, kjer je nizko zaupanje (rdeča), sta modela enaka in se zato tudi lepo prekrivata. Mesta z nizkim zaupanjem se nahajajo: eno na samem začetku, trije v domeni LRR-jev ter eno na koncu modela molekule. Na sliki je prikazana primerjava a) celotne molekule TLR (orientirana je tako, da se v ozadju vidi N-terminalni del (začetek molekule) ter v ospredju C-terminalni del oziroma domena TIR (konec molekule), b) domene levcinskih ponovitev ter c) domene TIR.

Na sliki 30/d so prikazani poravnani modeli TLR15\_1 (zelen), TLR15\_3 (vijoličen) in TLR15\_4 (rumen). Rezultati so zelo podobni kot pri primerjavi TLR15\_0 in TLR15\_1. Do neujemanja je prišlo na petih mestih, kjer je nizko zaupanje. Model molekule je na tej sliki obrnjen obratno, kot na sliki 30/a. Domena TIR oziroma C-terminalni konec molekule je v ozadju ter N-terminalni začetek molekule v ospredju. Za boljšo vidljivost so povečana mesta, kjer je nizka zaupljivost in so modeli različni. Vidimo lahko, da so vsi trije modeli molekul na teh mestih popolnoma različni.

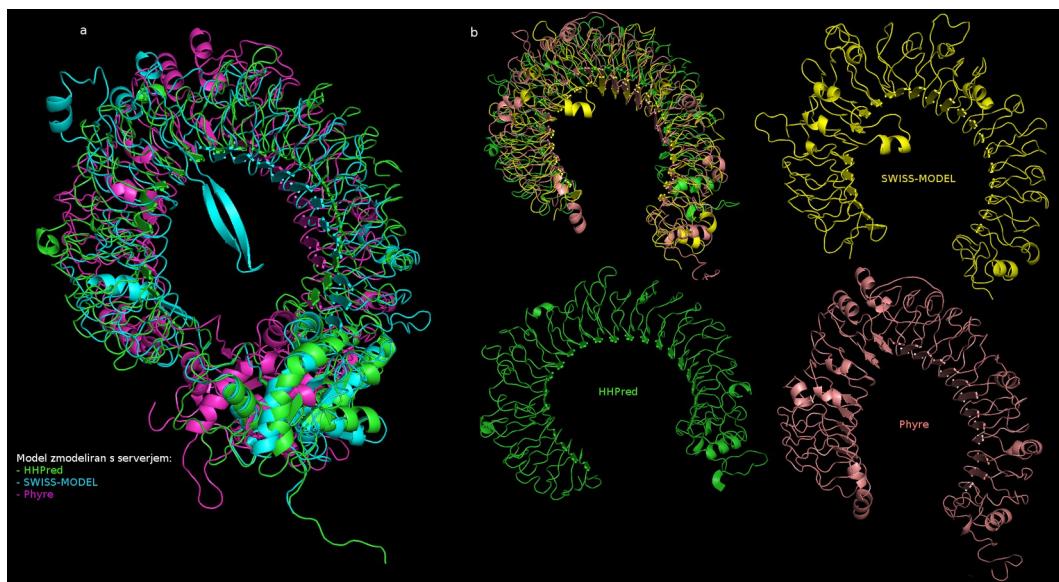


Slika 30: Primerjava modelov zmodeliranih s Phyre<sup>2</sup>

Tudi primerjava vseh modelov TLR15\_0, TLR15\_1 (zmodelirana posredno s strežnikom), TLR15\_3, TLR15\_4, TLR15\_5 ter TLR15\_6 (zmodelirani neposredno) je pokazala podobne rezultate kot prejšnji dve primerjavi – enak oziroma podoben model, razen na petih mestih z nizkim zaupanjem. Slika zaradi slabe preglednosti ni prikazana.

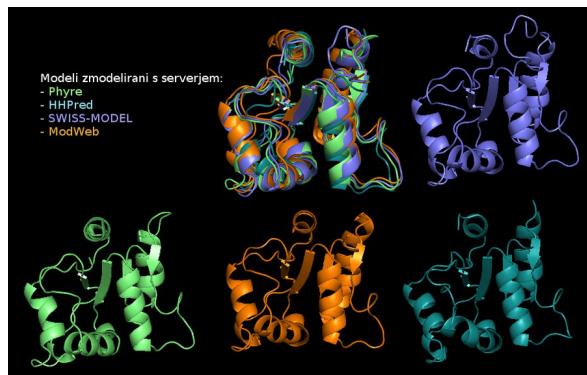
#### 4.7.3 Primerjava modelov zmodeliranih interaktivno s PMP (HHpred, SWISS-MODEL, ModWeb in Phyre<sup>2</sup>)

Slika 31 prikazuje a) primerjavo modelov TLR15 zmodeliranih s strežniki HHpred (zeleči), SWISS-MODEL (modra) in Phyre<sup>2</sup> (vijolična) ter b) primerjavo modelov domene



Slika 31: Primerjava modelov cele molekule ter modelov domene LRR-jev, zmodeliranih z različnimi strežniki

LRR-jev. Na sliki 32 je prikazana primerjava modelov domene TIR, kjer modra barva predstavlja model zmodeliran s strežnikom HHpred, zelena s strežnikom Phyre<sup>2</sup>, vijolična s strežnikom SWISS-MODEL ter oranžna s strežnikom ModWed.



Slika 32: Primerjava modelov domene TIR molekule TLR15

## 4.8 NAPOVED VEZAVNIH MEST ZA LIGANDE

### 4.8.1 Vezavna mesta napovedana z metodo COACH

Strežnik nam je ob napovedi vezavnih mest podal informacije o rezultatu (angl. *confidence score* – *C-score*), o številu šablon v gruči, o ligandih šablon ter o kompleksu ligand-protein, ter strukturo vseh ligandov v gruči. Napovedani ligandi so razvrščeni glede na rezultat (*C-score*). Rezultat ima lahko vrednost med 0 in 1, kjer večja vrednost pomeni bolj zanesljivo napoved. Na spletni strani strežnika je prikazanih 10 napovedi z najvišjim rezultatom ter podrobne informacije o napovedi.

#### 4.8.1.1 Vezavna mesta napovedana na modelu TLR15\_6

Napovedana vezavna mesta za model TLR15\_6 (ki smo ga zmodelirali s programom Phyre<sup>2</sup>) ter informacije o napovedi so prikazane na sliki 33. Slika je kopija spletnne strani z rezultati iz strežnika.

Click to view	Rank	C-score <sup>a</sup>	Cluster size <sup>b</sup>	Rep Temp <sup>c</sup>	Rep Lig <sup>d</sup>	Mult Temp <sup>e</sup>	Consensus Binding Residues
	1	0.03	3	227XB	227XB03	Download 436,438,455,456,457,458,481	
	2	0.03	3	3cyA	k-mer	N/A	392,416
	3	0.02	2	227XA	227XA04	Download 243,245,259,262,265,267,270,272,275,276,277,278,279,280,282,283,284,290,295,296,297,298,314,316,319,321,324,328,330,337,352	
	4	0.02	2	3C1Y	3C1Y03	Download 329,330,331,332,352,354,355,356,359	
	5	0.02	2	2281A	2281A01	Download 270,272,273,276,279,296,297,298,299,316,319,321,326,327,328,329,330,331,333,334,337,352,353,354,357	
	6	0.01	1	3A79B	3A79B02	Download 233,235,236,259,260,261,262,263,264,310	
	7	0.01	1	3A79B	3A79B04	Download 323,325,347	
	8	0.01	1	2A0ZA	2A0ZA10	Download 457,458,459,482,483,484,485	
	9	0.01	1	3cyB	Nuc Acid	Download 61,86,527	
	10	0.01	1	3A79A	3A79A02	Download 149,150,151,153,177,180	

Click [here](#) to download the residue-specific binding probability, which is estimated by SVM.  
Click [here](#) to download the predicted bound ligands and detailed prediction summary.  
Click [here](#) to download the templates clustering results.  
(a) C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.  
(b) Cluster size is the total number of templates in a cluster.  
(c) Rep Temp presents a predicted representative ligand-protein complex structure from a cluster.  
(d) Rep Lig is the bound ligand from the template. Click the ligand name to view its information in [the BioLIP database](#).  
(e) Mult Temp provides the structures of all ligands in a cluster.

Slika 33: Informacije o napovedi vezavnih mest na modelu TLR15 izračunanim s Phyre<sup>2</sup>

V preglednici 10 so za boljšo preglednost navedena imena šablon in ligandov, s pomočjo katerih je strežnik napovedal vezavna mesta. Šablone so obarvane glede na vrsto molekule. Za šablono je bila 9x uporabljena molekula TLR2, 6x TLR3, 2x TLR6 ter po 1x molekula TLR1, molekula NDH ter H-protein bakterije *Escherichia coli*.

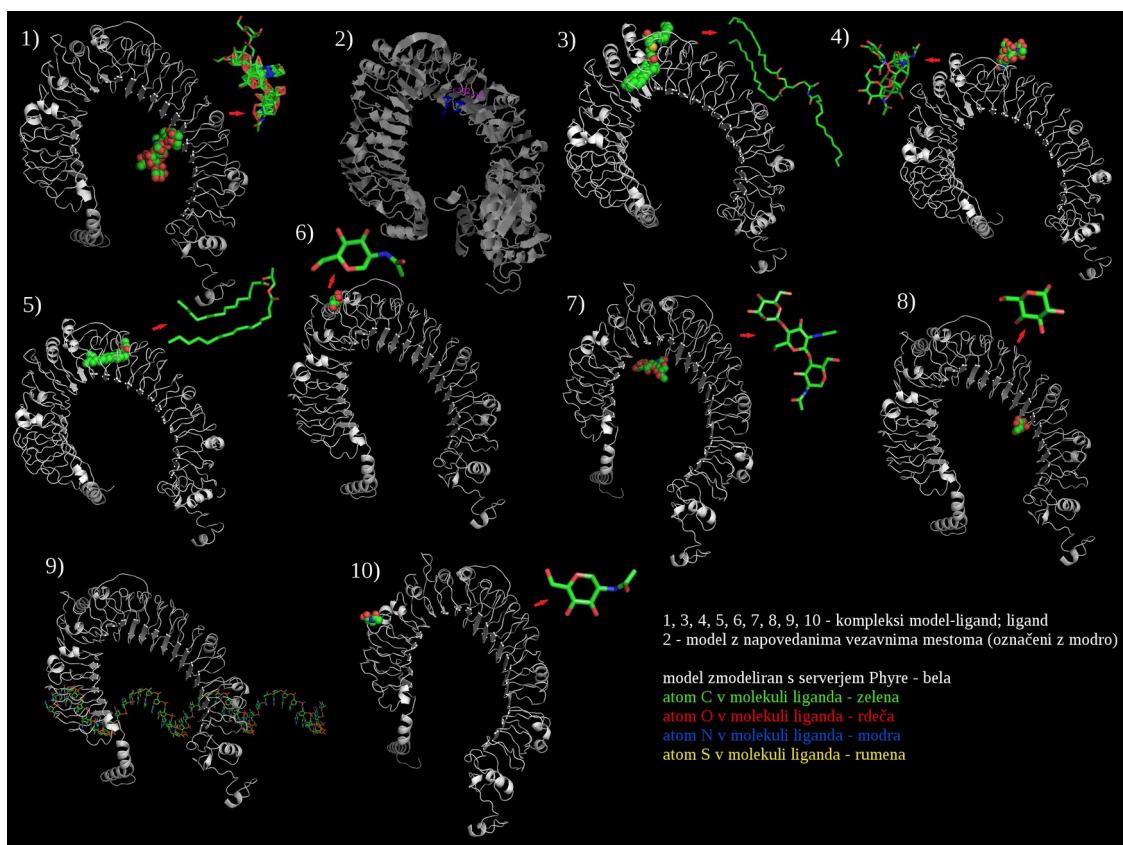
**Preglednica 10:** Šablone in ligandi uporabljeni za napoved vezavnih mest

Phyre <sup>2</sup>				I-TASSER			
PDB ID	Napoved	Šablona	Ligand	PDB ID	Napoved	Šablona	Ligand
2Z7XB	1	TLR1	sladkor*	2Z7XA	1	TLR2	sladkor*
3CIYA	2	TLR3	k-mer**	3A79A	2	TLR2	PAM2C
2Z7XA	3	TLR2	PAM3CS	2A0ZA	3	TLR3	NAG_NAG
3CIXC	4	TLR3	sladkor*	4G74A	4	NDH+Q	TRT
2Z81A	5	TLR2	PAM2C	2Z7XA	5	TLR2	K4
3A79B	6	TLR6	NAG	3AB9A	6	H-protein	Ca
3A79B	7	TLR6	NAG_NAG_BMA	3CIYA	7	TLR3	k-mer**
2A0ZA	8	TLR3	GLC	2Z81A	8	TLR2	NAG_NAG_MAN
3CIYB	9	TLR3	NUC_AC	2Z7XA	9	TLR2	PAMCSK4
3A79A	10	TLR2	NDG	2Z7XA	10	TLR2	NAG_NAG_MAN

\*posamezne molekule sladkorja so zbite v strukturo, ki ni prepoznavna

\*\*podani so le podatki o vezavnih mestih/ k-mer je n-zaporedje nukleinskih kislin

Najpogostejši ligandi vezani na napovedana vezavna mesta so različni sladkorji ter di-(PAM2) ali tri-acilirani (PAM3) lipopeptidi. Sladkorji se lahko pojavijo kot monomeri (NAG, NDG, GLC), dimeri (NAG\_NAG) ali polimeri (NAG\_NAG\_BMA, NAG\_NAG\_MAN). Pri nekaterih lipopeptidih je strežnik odstranil peptidni del (K4) ali



**Slika 34:** Kompleksi model (Phyre<sup>2</sup>)-ligand napovedni z metodo COACH

pa katero od lipidnih verig (PAM), zato so prikazani samo deli ligandov, ki so vezani na vezavna mesta. Kompleksi model-ligand so prikazani na sliki 34 ter oštrevlčeni od 1 do 10 v istem vrstnem redu kot so navedeni v stolpcu »Napoved« v preglednici 10 v sklopu Phyre<sup>2</sup>.

V primeru druge napovedi, modela z napovedanimi vezavnimi mesti ni bilo mogoče prenesti, zato je slika modela skopirana direktno iz spletne strani strežnika s podanimi rezultati.

#### 4.8.1.2 Vezavna mesta napovedana na modelu zmodeliranim s strežnikom I-TASSER

V primeru uporabe sekvence (molekule TLR15) kot vhodne datoteke za napoved vezavnih mest s strežnikom COACH, nam je najprej program I-TASSER izračunal 3D model, za ka-

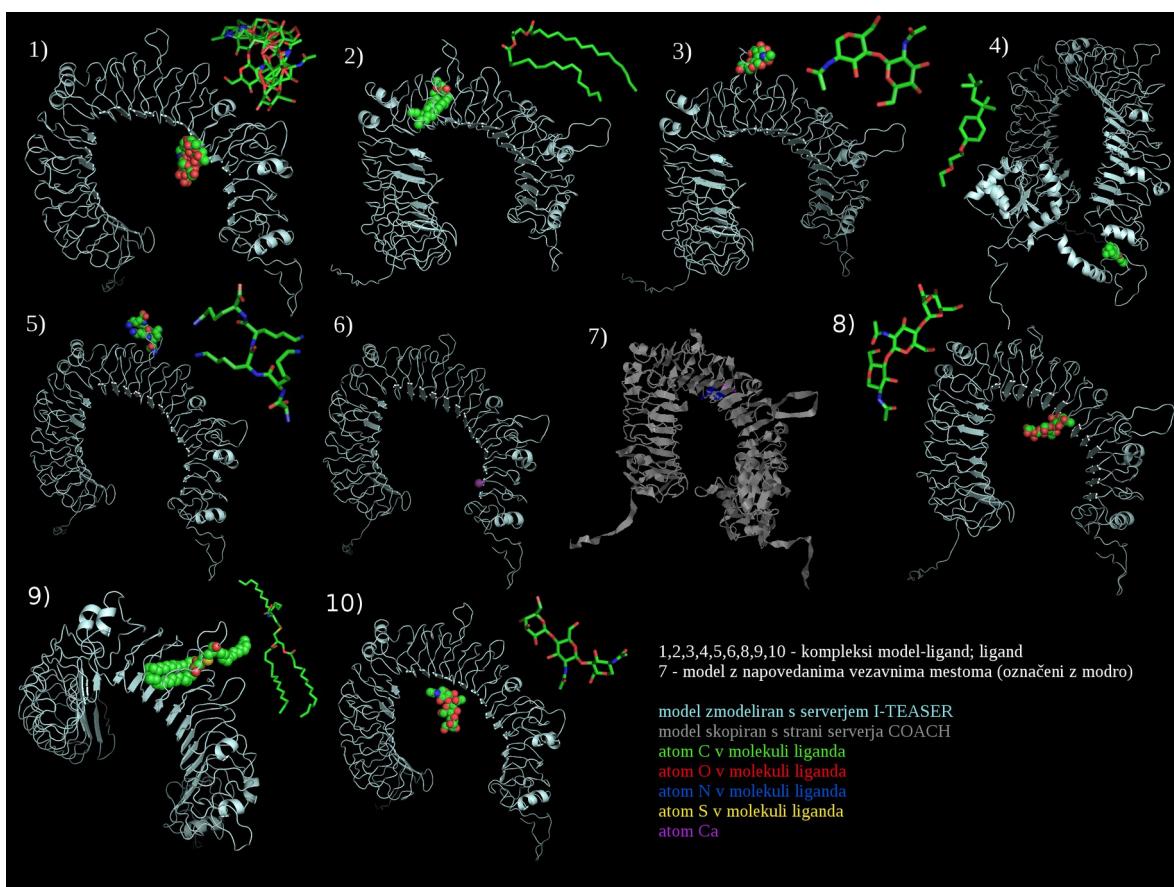
Click to view	Rank	C-score <sup>a</sup>	Cluster size <sup>b</sup>	Cluster Temp <sup>c</sup>	Rep Lig <sup>d</sup>	Mult Temp <sup>e</sup>	Consensus Binding Residues
●	1	0.04	4	<a href="#">2Z7XA03</a> <a href="#">Download</a>	418,436,437,438,439,457,458,459,460		
○	2	0.03	3	<a href="#">3A79A01</a> <a href="#">Download</a>	255,291,292,293,294,302,307,311,313,314,315,316,321,322,323,324,325,328,332,335,338,340,343,360		
○	3	0.02	2	<a href="#">2A0ZA06</a> <a href="#">Download</a>	349,351,354,355,356,357,358,359		
○	4	0.02	2	<a href="#">4g74A</a> <a href="#">TRT</a> <a href="#">Download</a>	678,681		
○	5	0.02	2	<a href="#">2Z7XA</a> <a href="#">2Z7XA00</a> <a href="#">Download</a>	348,349,350,351,352,353,354,355,356		
○	6	0.02	2	<a href="#">3ab9A</a> <a href="#">CA</a> <a href="#">Download</a>	571,572		
○	7	0.02	2	<a href="#">3c1yA</a> <a href="#">k-mer</a>	N/A	392,416	
○	8	0.02	2	<a href="#">2Z81A</a> <a href="#">2Z81A00</a> <a href="#">Download</a>	374,394,417,418,419,435,436,437,438,458,459		
○	9	0.01	1	<a href="#">2Z7XA</a> <a href="#">2Z7XA04</a> <a href="#">Download</a>	314,324,328,331,332,335,338,344,346,347,348,349,350,351,358,360,361,364,367,379,381,384,385		
○	10	0.01	1	<a href="#">2Z7XB</a> <a href="#">2Z7XB02</a> <a href="#">Download</a>	337,390,414		

Click [here](#) to download the residue-specific binding probability, which is estimated by SVM.  
Click [here](#) to download the predicted bound ligands and detailed prediction summary.  
Click [here](#) to download the templates clustering results.  
(a) C-score is the confidence score of the prediction. C-score ranges [0-1], where a higher score indicates a more reliable prediction.  
(b) Cluster size is the total number of templates in a cluster.  
(c) Rep Templ presents a predicted representative ligand-protein complex structure from a cluster.  
(d) Rep Lig is the bound ligand from the template. Click the ligand name to view its information in [the BioLIP database](#).  
(e) Mult Templ provides the structures of all ligands in a cluster.

Slika 35: Informacije o napovedi vezavnih mest na modelu TLR15 izračunanim s programom I-TASSER

terega je potem COACH napovedal vezavna mesta za ligande. Slika 35 prikazuje napovedana mesta ter dodatne informacije o napovedi. Imena šablon ter ligandov, ki jih je program uporabil za napoved, so navedena v preglednici 10 v prejšnjem razdelku. Slika 36 enako kot pri modelu TLR15\_6 tudi za ta model prikazuje kompleksne model-ligand, oštrevlčene v enakem vrstnem redu kot so razporejeni v preglednici 10 (v stolpcu »Napoved« v sklopu I-TASSER).

V sedmi napovedi, tako kot pri prejšnjem modelu pri drugi, modela ni bilo mogoče prenesti, zato je slika modela skopirana direktno iz spletne strani strežnika s podanimi rezultati.



Slika 36: Kompleksi model (I-TASSER)-ligand napovedni z metodo COACH

## 5 RAZPRAVA

Poznavanje 3D proteinske strukture je ključno za razumevanje biološke funkcije protein na molekularni ravni. Strukturo molekule lahko določimo eksperimentalno (z jedrsko magnetno resonanco ali rentgensko praškovno difrakcijo) ali pa določimo teoretično (z molekulskim modeliranjem) (Bordoli in Schwede, 2012). S pomočjo primerjalnega modeliranja molekul lahko napovemo funkcijo proteina ter identificiramo aktivna in vezavna mesta. S pomočjo modelov molekul pa lahko študiramo reakcije med proteini ter med proteinom in ligandom, razvijamo inhibitorje. Lahko tudi iščemo, razvijamo ter izboljšujemo ligande za določeno vezavno mesto, napovemo epitope za antigene ter simuliramo interakcije proteinov (Fiser, 2009). Za modeliranje katerikoli molekule proteina je potrebno poznati aminokislinsko sekvenco, ki je na voljo na spletu v prostih dostopnih podatkovnih bazah. NCBI in UniProt sta podatkovni bazi, ki sproti zbirata in arhivirata vse na novo poznane sekvence proteinov. Z njimi lahko z algoritmi za primerjavo poiščemo strukture molekul s podobnimi sekvincami, na podlagi katerih lahko s homolognim modeliranjem določimo strukturo želenemu proteinu.

Molekule TLR so receptorji prijenega imunskega sistema oziroma vzorčno prepoznavni receptorji, ki sodelujejo pri začetni fazi okužbe. So zelo razširjeni ter prisotni tako pri rastlinah kot pri živalih. Večina raziskav je osredotočenih na molekule TLR pri sesalcih, (predvsem mišje in človeške), kjer je bilo identificiranih 13 molekul TLR, pri ostalih vretenčarjih pa še nekateri dodatni (14-15, 18-23). Pri pticah je bilo identificiranih 10 molekul TLR (1La, 1Lb, 2a, 2b, 3, 4, 5, 7, 15, 21), od katerih naj bila TLR15 specifična samo za ptice (Brownlie in Allan, 2010). Struktura tega membranskega proteina še ni poznana, z nekaterimi eksperimentalnimi študijami so predvideli možne ligande, ki pa še niso potrjeni z računalniškimi modeli.

V nalogi sem s pomočjo bioinformacijskih orodij najprej poiskala vse znane sekvence molekule TLR15 ter jih primerjala med sabo. Z orodji za homologno modeliranje sem nato napovedala model molekule TLR15 ter možna vezavna mesta za ligande.

Eno leto po tem, ko je bil prvič posekveniran genom kokoši (*Gallus gallus*) je bil identificiran gen za protein TLR15 (Higgs in sod., 2006). Trenutno je v nepresežni podatkovni zbirki UniProt arhiviranih preko 60 sekvenc molekule TLR15 (preglednica 2), ki pripadajo 13 različnim vrstam ptic. Te številke se relativno hitro spreminja, saj se v podatkovno bazo sproti avtomatsko vnašajo vse na novo predložene sekvence. Največ vnosov (16) pripada vrstama *Gallus gallus* ter *Carpodacus mexicanus*, 13 vnosov pripada vrsti *Falco naumanni*, vse ostale vrste ptic pa imajo manj kot 4 vnoise. Vnesene sekvence so dveh različnih dolžin, približno 430 AK ter približno 870 AK. Slednje predstavljajo celotno molekulo, kraje sekvence pa glede na poravnave z orodjem BLAST predstavljajo del molekule z levcinskimi ponovitvami.

Nadalje smo primerjali podobnost med sekvincami molekule TLR15 pri različnih vrstah ptic. Najprej smo z orodjem za poravnavo več sekvenc Clustal Omega med seboj primerjali eno sekvenco posamezne vrste ptic, nato pa še dodatno z orodjem BLAST posamično primerjali sekvenco molekule TLR15 vrste *Gallus gallus* s sekvincami ostalih vrst.

Matrika podobnosti (slika 14), ki smo jo dobili s programom Clustal Omega nam prikazuje podobnost med sekvincami molekule TLR15 med posameznimi vrstami. Podobnosti med sekvenco TLR15 vrste *Gallus gallus* in sekvincami drugih vrst, ki smo jih dobili s progra-

mom Clustal Omega (rezultati so prikazani v matriki podobnosti) ter podobnosti dobljene s programom BLAST (preglednica 3) so enake. Podobnosti sovpadajo tudi s sorodnostjo med različnimi vrstami. Glede na filogenetsko drevo ptičev (Kimball in sod., 2013) so si med sabo najbolj sorodne vrste *Anas platyrhynchos* in *Anser anser*, (glede na matriko podobnosti imajo 94 % podobnost), ki pripadajo redu Anseriformes ter vrste rodu *Gallus* (medsebojna podobnost večja od 98 %) in vrsta *Meleagris gallopavo* (93 % podobnost z *Gallus*), ki pripadajo redu Galliformes, oba reda pa istemu nadredu (Galloanserae). Podobnost med redoma Anseriformes in Galliformes je odvisna od vrst, ki so predmet primerjave in je med 80 % in 83 %. Vse ostale vrste omenjene v nalogi pripadajo drugemu nadredu (Neoaves), znotraj katerega so si med seboj najbolj sorodne vrste *Carpodacus mexicanus*, *Petroica australis rakiura* ter *Taeniopygia guttata* (podobnost med slednjimi je med 84 % ter 88 %), ki pripadajo skupnemu redu Passeriformes. Slednji skupaj z družino Falconidae, kateremu pripadajo ptiči rodu *Falco*, pripada isti kladi (Eufalconimorphae). Podobnost med vrsto *Falco naumanni* ter vrstami iz redu Passeriformes je med 66 % in 69 %. Na filogenetskem drevesu ni prikazanega redu Procellariiformes, ki mu pripada vrsta *Oceanodroma leucorhoa*. Le ta ima največjo podobnost (82 %) z vrsto *Falco naumanni*. Najmanjša podobnost je po pričakovanjih med vrstami različnih nadredov (Galloanserae in Neoaves) in je med 60 % in 72 %.

Glede na podatke na spletu naj bi bilo do sedaj posekveniranih 9 genomov ptičjih vrst. 4 od teh imajo vneseno sekvenco molekule TLR15 v podatkovno zbirko UniProt, in sicer: *Anas platyrhynchos*, *Gallus gallus*, *Meleagris gallopavo* in *Taeniopygia guttata*. Vnosi sekvenc vseh ostalih vrst v podatkovni zbirki (preglednica 2) so pridobljeni ali s programi za napovedovanje genov ali pa so bili eksperimentalno določeni. Proteinske sekvene, ki so na voljo v podatkovni zbirki UniProt so namreč pridobljene s prevodom kodirajočih sekven, ki so predložene bazam nukleotidnih sekven (EMBL-Bank, GenBank in DDBJ) (The UniProt Consortium, 2008). Glede na to, da je bil najprej posekveniran genom vrste *Gallus gallus*, je tudi največ raziskav narejenih na genomu te vrste, tudi dosedanje študije o molekuli TLR15. Zato smo se odločili, da za nadaljnje delo izberemo sekvenco proteina TLR15 ptiča vrste *Gallus gallus*.

Preden smo se lotili modeliranja, smo za lažje razumevanje strukturne organizacije, s pomočjo programa LRRFinder, napovedali domene molekule. Domene vseh do takrat znanih molekul TLR so napovedali že Matsushima in sod. (2007), vendar z drugim programom (PFAM). Dobili so identične rezultate, le da so regijo med LRR3 in LRR4, ki smo jo mi označili kot neidentificirano (UN) (slika 16), označili kot regijo z dodatnimi AK ostanki, ki pripadajo LRR3 (slika 2). Glede na njihove napovedi, naj bi vsi TLR-ji, ki pripadajo družini TLR7 (TLR7, TLR8 in TLR9) imeli na koncu LRR15, še približno 40-60 dodatnih AK ostankov, ki naj bi predstavljalni zaščito hidrofobnega jedra. Glede na to, da se pri molekulji TLR15, regija z dodatnimi AK nahaja na drugem mestu (za LRR3 namesto za LRR15) in je poleg tega daljša (87 AK ostankov), obstaja možnost, da v tem primeru nima te funkcije. Posamezni LRR-ji modela molekule TLR15 so prikazani na sliki 30.

Kljub temu, da smo vedeli, da struktura TLR15 še ni eksperimentalno določena, smo s programom Protein Model Portal poskušali poiskati računalniške modele. Ker sekvenca ni pripadala nobenemu od modelov, ki so na voljo, smo z istim programom poskušali poiskati modele s podobnimi sekvencami. Večina že zmodeliranih molekul so bile domene TIR, ali pa deli LRR-jev (preglednica 5). Noben od modelov ne predstavlja celotne molekule, ali pa vsaj modela celotne ektodomene, ki je tarča za vezavo liganda (in ga potrebujemo za napo-

ved vezavnih mest) in vsi izračunani modeli so starejši od 2 let. Za pravilnost modela je zaželeno, da je izračunan čim kasneje, saj zelo hitro narašča število struktur, ki so lahko potencialne (še boljše) šablone za modeliranje.

S pomočjo bioinformacijskih orodij za homologno modeliranje smo izračunali več modelov molekule TLR15, ki smo jih primerjali med sabo. Kot sem že omenila v razdelku 2.3.1, je homologno modeliranje sestavljeno iz 4 osnovnih korakov: identifikacija šablone, poravnava med tarčno sekvenco in strukturo šablone, izračun 3D modela ter ocenjevanje pravilnosti modela. S portalom PMP je možno naenkrat zagnati modeliranje na različnih, glede na najnovejše raziskave CASP, najuspešnejših strežnikih, s čimer lahko prihranimo veliko časa. Vse rezultate, tudi povezave do strežnikov, po koncu modeliranja dobimo po elektronski pošti na želen naslov. Različni strežniki so nam posredovali različno količino podatkov oziroma informacij o modeliranju. Najmanj informacij nam je posredoval strežnik HHpred, ki nam je posredoval le eno .txt datoteko s koordinatami atomov ter podatki o šablonah. Največ podatkov pa nam je posredoval strežnik Phyre<sup>2</sup>, kjer smo s povezavo do rezultatov na strežniku imeli vpogled v informacije o modeliranju (naštete v razdelku 4.5.4). Količina podatkov, ki nam jih posredujejo strežniki so odvisni od metod in algoritmov, ki se uporabijo za izvedbo posameznih korakov modeliranja ter interpretacijo dobljenih rezultatov s strani posameznega strežnika, za razumevanje katerih je potrebno imeti dodatno računalniško znanje. V razdelku o pregledu objav so razen za strežnik Phyre<sup>2</sup>, s katerim smo izračunali končni model, opisani le osnovni principi modeliranja za posamezne strežnike.

Vsek strežnik nam je v obliki PDB ID posredoval informacije o strukturah molekul, ki jih je izbral kot šablone za modeliranje. Glavni vir za iskanje šablon za homologno modeliranje je podatkovna baza PDB, ki vsebuje vse eksperimentalno določene strukture. Šablone v uvodnem koraku modeliranja strežniki določijo z metodami in algoritmi za poravnavo sekvenc ter za modeliranje izberejo tiste, ki imajo največjo sekvenčno podobnost s tarčno molekulo (v nadaljevanju podobnost). Informacije o podobnosti v % so prikazane pri strežnikih SWISS-MODEL, ModWeb in Phyre<sup>2</sup>, medtem ko nam je strežnik HHpred posredoval le PDB ID šablon brez dodatnih informacij o modeliranju. Za modeliranje molekule TLR15 je bilo izbranih pri različnih strežnikih različno število šablon (prikoga A) in od izbrane šablone je odvisno kateri del regije proteina bo zmodeliran. Strežnik HHpred je en modela celotne molekule izračunal s pomočjo 9 šablon (preglednica 6). Strežnika SWISS-MODEL ter ModWeb sta izračunala več modelov, ki zajemajo različne regije proteina (domeno TIR, ektodomeno, posamezne levcinske ponovitve). Za vsak izračunan model sta uporabila eno šablon. Strežnik Phyre<sup>2</sup> je izračunal en model na podlagi več šablon. Princip modeliranja s Phyre<sup>2</sup> se razlikuje glede na izbrani način modeliranja (»Intensive« ali »Normal«) – razlike so opisane v razdelku 2.4.4.4.

Pri vseh strežnikih je bila za izračun modela domene TIR kot šablona uporabljena struktura domene TIR molekule TLR1 (PDB ID: 1FYV\_A). Slednja je glede na podobnost določeno pri strežnikih SWISS-MODEL (47 %), ModWeb (50 %) ter Phyre<sup>2</sup> (47 %), najbolj homologna domeni TIR molekule TLR15.

Precej manjše podobnosti so bile določene za ektodomeno ter celo molekulo. Strežnik SWISS-MODEL je edini, ki nam je posebej posredoval model ektodomene, za modeliranje katere je izbral strukturo molekule TLR4 (PDB ID: 4G81\_B) z 20 % podobnostjo.

Model celotne molekule so izračunali strežniki SWISS-MODEL, Phyre<sup>2</sup> ter HHpred. Prvi je kot šablon za modeliranje uporabil strukturo molekule TLR5 (PDB ID: 3J0A\_A), osta-

la dva sta uporabila več šablon. Izmed vseh šablon, izbranih s strani strežnika Phyre<sup>2</sup>, priпадa največja podobnost (22 %) molekuli TLR8 (PDB ID: 3W3J\_B), ki pa pokriva le ektdomeno proteina. Izmed vseh izračunanih modelov je Phyre<sup>2</sup> v načinu »Normal« za najboljšega izbral tistega, ki je bil izračunan s pomočjo šablone TLR5 (PDB ID: 3j0A\_A), ki ji pripada 18 % podobnost. Za model izračunan v načinu »Intensive« je strežnik uporabil vse izbrane šablone, saj za izračun končnega modela zajame podatke več modelov izračunanih s pomočjo šablon z največjo podobnostjo. ModWeb ni izračunal niti modela celotne molekule, niti modela ektdomene, ampak le določene regije demene LRR-jev, ki pa za našo nalogu niso relevantne.

Glede na rezultate torej lahko povzamemo, da je domena TIR molekule TLR15 najbolj homologna domeni TIR molekule TLR1 ter, da je ektdomena najbolj homologna molekuli TLR8. Naj omenim, da je celoten model molekule TLR (ektdomena + domena TIR) eksperimentalno določen le za molekulo TLR5, za vse ostale molekule TLR so določene le strukture ektdomene. To je verjetno vzrok, da je bila ta molekula izbrana kot šablon za modeliranje celotne molekule pri vseh strežnikih, razen pri ModWeb (ki modela celotne molekule ni izračunal). Glede na to, da je na voljo več struktur ektdomene, kot pa celotne molekule, bi bilo bolj optimalno, če bi zagnali modeliranje ektdomene same in potem primerjali homologije.

Glede na % podobnosti, lahko homologijo posameznih domen ali pa celotne molekule, razvrstimo v cone sekvenčne podobnosti in tako približno napovemo kvaliteto modela. V našem primeru so nam oceno kvalitete modela, z izjemo strežnika HHpred, izračunali strežnički sami. Ker vsak strežnik predstavi oceno modela na drugačen način je kvaliteto med različnimi strežniki težko primerjati med seboj. V primeru domene TIR, kjer je podobnost v primerjavi z ektdomeno ter celotno molekulo največja (več kot 45 %), je bil po pričakovanjih izračunan model najkvalitetnejši, kar potrjujejo tudi ocene strežnikov SWISS-MODEL (preglednica 7, slika 21) ter Phyre<sup>2</sup> (slika 23). V primeru strežnika ModWeb ne moremo narediti primerjave, ker ni izračunanega modela ektdomene ter celotnega modela molekule, v primeru strežnika HHpred pa ocena kvalitete ni bila podana. S pomočjo strežnika SWISS-MODEL lahko primerjamo tudi kvaliteto med modelom ektdomene ter modelom celotne molekule. V primeru ektdomene, ki ji priprada večja podobnost (20 %) je tudi kvaliteta modela boljša, kot v primeru modela celotne molekule (podobnost 19 %) (preglednica 7, slika 21). Pri strežniku Phyre<sup>2</sup>, ki nam je kot rezultat podal le en končni model celotne molekule, te primerjave ne moremo narediti. Glede na zaupanje (barvna lestvica) (slika 23), lahko predvidimo, da je del modela, kjer se nahaja domena TIR (regija 705-845), kvalitetnejši od celotnega modela oziroma dela modela, kjer se nahaja ektdomena, saj prvi ne vsebuje regij zmodeliranih *ab initio*, drugi pa. Na spletni strani strežnika Phyre<sup>2</sup> navajajo, da je z njihovim postopkom možno zaznati tudi zelo oddaljene homologije (s podobnostjo do 15 %) in zmodelirati pravilne modele, kljub majhni podobnosti. Modeliranje smo s Phyre<sup>2</sup> ponovili 7 krat – 2 krat posredno preko portala PMP v načinu »Intensive«, 1 krat neposredno s strežnikom v načinu »Normal« ter 4 krat neposredno v načinu »Intensive«. Modele smo med sabo primerjali.

Vsi izračunani modeli (s strežnikom HHpred, SWISS-MODEL, Phyre<sup>2</sup>) celotne molekule, imajo po pričakovanju podobne strukture. Ektodomena je sestavljena iz 19-20 LRR-jev (preglednica 4, slika 29) ter ima solenoidno strukturo v obliki podkve, pri kateri so hidrofobni ostanki usmerjeni v notranjost molekule. Domena TIR ima značilno že poznano strukturo, kar prikazujejo izračunani modeli, ki smo jih primerjali med sabo (slika 32).

Transmembranska domena naj bi bila glede na rezultate strežnika Phyre<sup>2</sup>, ki v načinu »Normal« napove tudi topologijo molekule, dolga 30 AK ostankov in se v proteinu nahaja na mestu 655 – 685. Drugačno dolžino je napovedal program TollML in sicer, da je domena dolga 13 AK in se nahaja na mestu 663 – 676. Za natančnejšo določitev transmembranske domene, bi bilo optimalno uporabiti še kakšen dodaten strežnik za napoved transmembranskih domen. Model te domene je bil izračunan le v primeru celotne molekule in glede na oceno kvalitete modela, izračunanega s Phyre<sup>2</sup>, ima ta regija modela nizko zaupanje in je zmodelirana *ab initio*, kar pomeni da model tega dela strukture ni zanesljivo pravilen.

Primerjave modelov, ki smo jih naredili s pomočjo programa Pymol, so pokazale, da so vsi modeli enake oblike. Zaradi preglednosti in količine podatkov nismo primerjali vsakega modela s vsakim. Smiselno smo poskušali prikazati grobe razlike med modeli izračunanimi z istim strežnikom (Phyre<sup>2</sup>) ter med posameznimi strežniki. Za določitev vezavnega mesta za ligande ter samih ligandov, je pomembna struktura ektodomene. Glede na vse dosedanje raziskave o TLR-jih in njihovih ligandih, se vezavno mesto nahaja na tem delu molekule. Primerjava modelov izračunanih s strežnikom Phyre<sup>2</sup>, je po pričakovanjih pokazala, da se ti modeli, razen na mestih zmodeliranih *ab initio*, prekrivajo. Modeliranje *ab initio* ne temelji na homologiji, zato se te strukture med seboj razlikujejo. Strežnik jih označi kot regije z nizkim zaupanjem oziroma verjetnostjo, da ta del modela ni pravilen. Ker na teh mestih ni zaznane homologije, to pomeni da je specifično za to molekulo, zaradi česar ima to mesto lahko kakšno specifično funkcijo. V primeru, da molekula veže ligand, ki ni specifičen za ostale molekule TLR, so ta mesta lahko potencialna mesta za vezavo tega neznanega liganda, vendar bi za potrditev bile potrebne nadaljnje študije.

Razlika med modeli, izračunanimi z različnimi strežniki, je po pričakovanjih večja, kot med modeli, izračunanimi z enim strežnikom. To je zaradi uporabe različnih metod in tehnik, ki vplivajo na različne rezultate že pri vsakem posameznem koraku modeliranja. Tako so različni modeli posledica različnih homologij zaznanih z različnimi metodami, različnih večsekvenčnih poravnav homolognih sekvenc ter na koncu tudi različnih metod za izračun modelov. Kljub temu je v osnovi oblika strukture enaka, čeprav prihaja tudi do razlik medstrukturami posameznimi LRR-jev zmodeliranih s posameznimi strežniki. Primerjava modelov domen LRR-jev še lepše prikazuje obliko molekule, ki je pri vseh treh modelih enaka, večje razlike so opazne na začetku ter koncu modela ektodomene, kjer ni zaznanih homologij.

Če bi želeli zmodelirati bolj pravilen model bi bilo zaželeno, da se zmodelira vsaka domena posebej s trenutno najboljšimi stežniki na spletu (ocenjenim glede na CASP). Raziskave CASP potekajo vsaki dve leti, kar pomeni da ima vsake toliko časa katerikoli strežnik možnosti da postane najboljši. Strežniki so vedno boljši, napovedujejo čedalje pravilnejše modele, jih je tudi čedalje več in je med njimi čedalje večja konkurenca. Seveda je kvaliteta strežnika odvisna od kvalitete modelov, ti pa od vseh korakov v homolognem modeliranju. Dober strežnik torej uporablja optimalne algoritme, ki čim bolj pravilno in uspešno izpeljejo vsak posamezni korak modeliranja. Najboljši strežnik je trenutno I-TASSER, pri katerem je bila v času raziskovalnega dela potrebna registracija. Ker smo za modeliranje uporabili prosto dostopne strežnike, tega nismo uporabili neposredno za napoved modela. Smo pa po naključju prišli do napovedanega modela s strani tega strežnika z uporabo programa za napoved vezavnih mest za ligande (v nadaljevanju).

Vezavna mesta za ligande ter ligande smo napovedali s pomočjo strežnika COACH, ki za napoved prav tako uporablja podobnost oziroma homologije in je trenutno (glede na raziskave CAMEO) najbolj uspešen na področju napovedi vezavnih mest. Strežnik ima možnost, da kot vhodno datoteko sprejme ali že znano strukturo ali pa sekvenco, ki ji nato s pomočjo strežnika I-TASSER določi strukturo. Slednji je glede na raziskave CASP trenutno na prvem mestu, glede na uspešnost homolognega modeliranja. Mi smo zaganali napoved vezavnih mest z eno od naših struktur (ektodomeno modela TLR15\_6) ter s sekvenco, ki smo jo uporabili za modeliranje (sekvenco ektodomene). Tako smo dobili napovedi za naš model ter za model izračunan s strežnikom I-TASSER. Rezultati so bili za oba modela podobni (preglednica 10). Oba modela sta kot šablono za napoved v največ primerih izbrala molekulo TLR2. Ta tvori dimer v kompleksu z molekulom TLR6 ob vezavi diaciliranega lipopeptida - PAM2CSK4 (Kang in sod., 2009) ter v kompleksu z molekulom TLR1 ob vezavi triaciliranega lipopeptida – PAM3CSK4 (Jin in sod., 2007). Pri obeh modelih je bil v dveh primerih za ligand izbran lipopeptid, v ostalih primerih pa sladkorji, ki so posledica glikozilacije. V kompleksu se na vezavno mesto molekule TLR2 vežeta estersko vezana lipida aciliranega lipopitpida. Tudi v primeru modela molekule TLR15 je ne glede na to ali je ligand PAM2CSK ali PAM3CSK, na napovedano vezavno mesto vezan ta del liganda. Strežnik je napovedal tudi potencialna vezavna mesta za nukleinske kisline, kjer je za šabloni izbral molekulo TLR3. Glede na to, da je kot šablono bila večkrat izbrana molekula TLR2, to pomeni, da je večja verjetnost, da vezavnim mestom pripada večja homologija za aciliran lipopeptid. Strežnik je napovedal še vezavna mesta na podlagi šablon, ki niso molekule TLR ter vežejo ligande, ki za molekule TLR na podlagi dosedanjih študij niso specifični, zato sklepamo, da to niso vezavna mesta ter ligandi molekule TLR15.

Vsa štiri vezavna mesta (za acilirane lipopeptide) napovedana s strežnikom se nahajajo na približno enaki regiji proteina (243 – 352, 270-357, 255-360, 348-356 ). Tore lahko pričakujemo, da se vezavno mesto nahaja na mestu od 243. do 357. aminokislinskega ostanka. Prav na tem delu molekule je območje, ki ima glede na model izračunan s strežnikom Phyre<sup>2</sup>, nizko zaupanje iz česar lahko sklepamo, da TLR15 veže ligand na poseben način. Ker pri ptičih ni molekule TLR6, lahko pričakujemo da se diacilirian lipopeptid veže na drugačen način kot pri sesalcih in da imajo namesto TLR6 drug TLR za vezavo teh ligandov. Zagotovo lahko trdimo da se ob vezavi diaciliranega peptida ne tvori dimer TLR2-TLR6, torej se ta ligand pri ptičih veže na drugačen način. TLR15 je glede na rezultate Owen in sod. (2013) ter računalniške napovedi v naši nalogi potencialni receptor za vezavo diaciliranega lipopeptida.

Naj še enkrat poudarim, da so vsi napovedani modeli izračunani na podlagi homologije, torej že znanih in dokazanih eksperimentalnih struktur. S homolognim modeliranjem lahko napovemo kvalitetne teoretične modele molekul, ne moremo pa napovedati novih struktur ter vezavnih mest za ligande. Tudi v tem primeru je sam model izračunan na podlagi že znanih struktur ter že znanih vezavnih mest, ki imajo z molekulom TLR15 največjo podobnost, torej je to le teoretična napoved. Če bi želeli dobiti popolnoma pravilno strukturo molekule ter vezanega liganda, bi jo morali določiti še eksperimentalno.

## 6 SKLEPI

- Trenutno je v nepresežni podatkovni zbirki UniProt arhiviranih dobrih 60 sekvenc molekule TLR15, ki pripadajo 13 različnim vrstam ptičev
- Podobnosti med sekvencami molekule TLR15 sovpadajo s sorodnostjo med različnimi vrstami. Bolj so si sorodne vrste med sabo, večja je podobnost med sekvenca mi.
- Domena TIR molekule TLR15 je najbolj homolgona domeni TIR molekule TLR1, ektodomena je najbolj homologna molekuli TLR8, celotna molekule pa molekuli TLR5, ki ima edina eksperimentalno določeno strukturo
- Kvaliteta modela molekule je boljša, če je sekvenčna podobnost med tarčno moleku ter šablono večja. Kvalitete modelov, izračunane z različnimi strežniki ne moramo primerjati med sabo.
- Model molekule TLR15 ima značilno solenoidno strukturo, z regijami specifičnimi za ta protein. Te regije modela s homolognim modeliranjem ni bilo mogoče napovedati, zato obstaja verjetnost, da ta del modela ni pravilno napovedan.
- Regije, specifične za molekulo TLR15, so potencialna vezavna mesta proteina. Glede na homologijo vezavnih mest, je potencialni ligand molekule TLR15 diaciliran lipopeptid.

## 7 POVZETEK

Bogastvo proteinskih struktur, ki so na voljo, nam omogoča neprimerljivo priložnost za študij in boljše razumevanje načel o zvijanju proteinov in evoluciji njihove strukture. Sodobne eksperimentalne metode za določanje proteinske strukture (jedrska magnetna resonanca in rentgenska praškovna difrakcija) lahko razrešijo le majhen delež proteinov, predvsem zaradi časovnih in tehnoloških omejitev. Pri tem nam lahko pomagajo računalniške metode za napovedovanje proteinskih struktur, ki se konstantno razvijajo. Ena takih metod je homologno modeliranje, ki temelji na dejstvu, da imajo sorodne proteinske sekvence podobno 3D strukturo.

Homologno modeliranje se uporablja za napoved 3D strurnih modelov proteinov, ki jih izračuna s pomočjo sekvenčne poravnave med tarčo in šablono z že poznano strukturo. Trenutno je na voljo kar nekaj strežnikov za napoved strukture proteinov, katerih uspešnost se vsake dve leti preverja z raziskavami CASP. Glede na zadnje rezultate najpravilnejše modele generira strežnik I-TASER, med zelo uspešne pa sodita tudi Phyre<sup>2</sup> ter HHpred, s pomočjo katerih smo v nalogi napovedali strukturo molekule TLR15.

Receptorji podobni Toll-u oziroma receptorji TLR so membranski proteini, ki sprožijo prirojen imunski odziv po tem ko se nanje vežejo različne spojine patogenih organizmov. Te spojine imenujemo ligandi in so lahko bakterijskega, virusnega, glivnega ali praživalskega izvora. Strukutra molekule TLR je sestavljena iz treh domen: N-terminalne domene, ki se nahaja izven celice ter prepoznavata ligande, transmembranske domene – heliks ter C-terminalne, signalne domene, ki se nahaja znotraj celice. TLR15 je značilen le za ptiče in se molekulsko razlikuje od ostalih znanih molekul TLR. Ker večina raziskav poteka na molekulah TLR pri sesalcih, struktura TLR15 še ni določena, poznani pa tudi niso ligandi tega receptorja.

Namen dela je bil s pomočjo homolognega modeliranja napovedati 3D strukturo molekule TLR15 in vezana mesta za ligande ter preveriti potencialne ligande za ta protein.

V nalogi sem v uvodnih poglavjih pri pregledu objav najprej na kratko opisala klasifikacijo proteinov, ki je pomembna pri interpretaciji njihove strukture. Nadalje sem opisala vlogo receptorja TLR v imunskeistem ter navedla strurne lastnosti te molekule in znane ligande, ki se vežejo na vezavna mesta domene levcinskih ponovitev. Za napoved strukture molekule s homolognim modeliranjem je pomembno, da so na voljo strukture sorodnih molekul s čim večjo sekvenčno podobnostjo, ki ima velik vpliv na kvaliteto končnega modela. Za boljše razumevanje principa homolognega modeliranja ter njegovega pomena v bioinformatiki in naravoslovju, sem opisala nekatere bioinformačne metode in navedla bioinformacijska orodja, s pomočjo katerih je možno napovedati kvalitetne modele ter določiti vezavna mesta za ligande.

V nalogi smo najprej s pomočjo bioinformacijskih orodij poiskali (UniProt, NCBI) in primerjali (BLAST, Clustal Omega) znane sekvence molekule TLR15 ter eno izmed njih (*Gallus gallus* - C4PBP0) uporabili za modeliranje. Preden smo zagnali modeliranje, smo s pomočjo portala PMP preverili, če na spletu že obstaja model molekule. Portal ni našel nobenega modela, ki bi ustrezal sekvenci proteina TLR15, zato smo uporabili interaktivno modeliranje portala, kjer smo za izračun modela posredno uporabili strežnike SWISS-MODEL, HHpred, ModWeb ter Phyre<sup>2</sup>. Slednjega smo uporabili tudi neposredno za napoved

modela. Napovedane modele smo primerjali, končni model pa analizirali in napovedali vezavna mesta ter potencialne ligande.

S pomočjo orodij za iskanje in primerjavo sekvenc smo v podatkovni zbirki UniProt dobili več kot 60 sekvenc molekule TLR različnih vrst ptic in pokazali, da je podobnost teh sekvenca odvisna od sorodnosti med vrstami. Največkrat je v podatkovno zbirko vnesena sekvenca TLR15 vrste ptiča *Gallus gallus*, katero smo uporabili za modeliranje. S pomočjo informacij različnih strežnikov o modeliranju molekule, smo primerjali sekvenčno podobnost posameznih domen molekule TLR15 z ostalimi molekulami TLR ter ugotovili, da je domena TIR najbolj homologna domeni TIR molekule TLR1, domena levcinskih ponovitev pa molekuli TLR8. S pomočjo strežnikov, ki so nam posredovali informacije o kvaliteti modela smo pokazali, da je kvaliteta izračunanega modela boljša, če je sekvenčna podobnost med tarčno molekulo in šablono večja. Iz dobljenih modelov lahko sklepamo, da ima molekula TLR15 značilno solenoidno strukturo z regijami, ki so specifičnimi za ta protein. Glede na rezultate napovedi vezavnih mest lahko sklepamo, da so te specifične regije potencialna vezavna mesta za ligand ter da je potencialni ligand molekule diaciliran lipopeptid.

Naloga je lep primer uporabe bioinformatike v primeru napovedi neznane strukture proteina na podlagi homologije. Metode za poravnavo sekvenca ter homologno modeliranje so ene izmed tistih, ki bodo v prihodnje s svojo dostopnostjo in čedalje večjo zmogljivostjo ter lažjim rokovanjem, veliko pripomogle k raziskavam v različnih naravoslovnih disciplinah.

## 8 VIRI

- Altschul S. F. 1991. Amino Acid Substitution Matrices from an Information Theoretic Perspective. *Journal of Molecular Biology*, 219: 555-565
- Altschul S. F., Gish W., Miller W., Myers E., Lipman D. J. 1990. Basic local alignment search tool. *Journal of Molecular Biology*, 215: 403-410
- Andreeva A. 2012. Classification of Proteins: Available Structural Space for Molecular Modeling. V: Methods in Molecular Biology. Vol. 857. Homology Modeling: Methods and Protocols. Orry J.W. A., Abagyan R. (eds.). New Jersey, Humana Press: 1-31
- Bella J., Hindle K. L., McEwan P. A., Lovell S. C. 2008. The leucine-rich repeat structure. *Cell and Molecular Life Sciences*, 65: 2307-2333
- Biegert A., Söding J. 2009. Sequence context-specific profiles for homology searching. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 10: 3770-3775
- Bordoli L., Schwede T. 2012. Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal. V: Methods in Molecular Biology. Vol. 857. Homology Modeling: Methods and Protocols. Orry J.W. A., Abagyan R. (eds.). New Jersey, Humana Press: 107-136
- Botos I., Segal D. M., Davies D. R. 2011. The Structural Biology of Toll-like Receptors. *Structure*, 19: 447-459
- Brownlie R., Allan B. 2011. Avian toll-like receptors. *Cell and tissue research*, 343: 121-130
- Chao K. M., Zhang L. 2009. Sequence Comparison/ Theory and Methods. Vol. 7. London, Springer London: 230 str.
- Eswar N., John B., Mirkovic N., Fiser A., Ilyin V. A., Pieper U., Stuart A. C., Marti-Renom M. A., Madhusudhan M. S., Yerkovich B., Sali A. 2003. Tools for comparative protein structure modeling and analysis. *Nucleic Acids Research*, 31, 13: 3375-3380
- Fekonja O., Avbelj M., Jerala R. 2012. Suppression of TLR Signaling by Targeting TIR domain-Containing Proteins. *Current Protein and Peptide Science*, 13: 776-788
- Fiser A. 2009. Comparative Protein Structure Modeling. V: From Protein Structure to Function with Bioinformatics. Rigden D.J. (ed.). Springer: 57-90
- Gay N. J., Gangloff M. 2007. Structure and function of Toll Receptors and Their Ligands. *Annual Review of Biochemistry*, 76: 141-165
- Gong J., Wei T., Zhang N., Jamitzky F., Heckl W. M., Rössle S. C., Star R. W. 2010. Toll-ML: a database of toll-like receptor structural motifs. *Journal of Molecular Modeling*, 16: 1283-1289
- Haas J., Roth S., Arnold K., Kiefer F., Schmidt T., Bordoli L., Schwede T. 2013. The Protein Model Portal – a comprehensive resource for protein structure and model information. *Database*, 2013, doi:10.1093/database/bat031:

- Henikoff S., Henikoff J. G. 1992. Amino acid substitution matrices from protein blocks. Proceedings of the National Academy of Sciences of the United States of America, 89: 10915-10919
- Higgs R., Cormican P., Cahalane S., Allan B., Lloyd A. T., Meade K., James T., Lynn D. J., Babiuk L. A., O'Farrelly C. 2006. Induction of a novel chicken Toll-like receptor following *Salmonella enterica* serovar Typhimurium infection. Infection & immunity, 74, 3: 1692-1698
- Hoeppner M., Ostell J. 2013. The NCBI Handbook. 2<sup>nd</sup> ed. Bethesda, National Center for Biotechnology Information.  
<http://www.ncbi.nlm.nih.gov/books/NBK143764/> (jan. 2014)
- Isaev A. 2006. Introduction to Mathematical Methods in Bioinformatics. Berlin, Springer: 294 str.
- Jin M. S., Kim S. E., Heo. J. Y., Lee M. E., Kim H. M. Paik S. G., Lee H., Lee J. O. 2007. Crystal structure of the TLR1-TLR2 Heterodimer Induced by Binding of a Tri-Acylated Lipopeptide. Cell, 130, 6: 1071-1082
- Kang J. Y., Nan X., Jin M. S., Youn S. K., Ryu Y. H., Mah S., Han S. H., Lee H., Paik S. G., Lee J. O. 2009. Recognition of Lipopeptide Patterns by Toll-like Receptor 2-Toll-like Receptor 6 Heterodimer. Immunity, 31, 6: 873-884
- Kelley L. A., Jefferys B. 2011. Phyre<sup>2</sup>. Structural Bioinformatics Group, Imperial College London.  
<http://www.sbg.bio.ic.ac.uk/phyre2/html/page.cgi?id=index> (nov. 2013)
- Kimball R., Wang N., Heimer-McGinn V., Ferguson C., Braun W. L. 2013. Identifying localized biases in large datasets: A case study using the avian tree of life. Molecular phylogenetics and evolution, 69, 3: 1021-1032
- Kitchovitch S., Song Y., van der Wath R., Lio P. 2009. Substitution Matrices and Mutual Information Approaches to Modeling Evolution. V: Learning and Intelligent Optimization. Third International Conference, LION 3, Trento, Italy, 14. -18. januar 2009. Stützle T. (ur.). Springer Berlin Heidelberg: 259-272
- Kobe B., Kajava A. Y. 2001. The leucine-rich repeat as a protein recognition motif. Current Opinion in Structural Biology, 11, 6: 725-732
- Kryshtafovych A., Krzysztof F. 2008. Protein Structure Prediction and Model Quality Assessment. Drug Discovery Today, 14, 7/8: 386-393
- Kumar H., Kawai T., Akira S. 2009. Toll-like receptors and innate immunity. Biochemical and biophysical research communications, 388, 4: 621-625
- Laurie A. T. R., Jackson R. M. 2006. Methods for the Prediction of Protein-Ligand Binding Sites for Structure-Based Drug Design and Virtual Ligand Screening. Current Protein and Peptide Science, 7, 5: 395-406
- Lester S. N., Li K. 2013. Toll-Like Receptors in Antiviral Innate Immunity. Journal of Molecular Biology. 426: 1246-1264
- Lopez R., Robinson S., Kibria A., Harte N., Patel G., Harper R., Quevillon E., Silventoinen V., Kallio K., Jokinen P. 2002. The European Bioinformatics Institute web site: a new view. Bioinformatics, 19, 4: 546-547

- Lopez R., Silventoinen V., Robinson S., Kibra A., Warren G. 2003. WU-Blast2 server at the European Bioinformatics Institute. Nucleic Acid Research, 31, 13: 3795-3798
- Lorenza B., Schwede T. 2012. Automated Protein Structure Modeling with SWISS-MODEL Workspace and the Protein Model Portal. V: Methods in Molecular Biology. Homology Modeling: Methods and Protocols. Vol. 857. Orry J.W. A., Abagyan R. (eds.). New Jersey, Humana Press: 107-136
- Matsushima N., Tanaka T., Enkhbayar P., Mikami T., Taga M., Yamada K., Kuroki Y. 2007. Comparative sequence analysis of leucine-rich repeats (LRRS) within vertebrate toll-like receptors. BMC Genomics, 8: doi:10.1186/1471-2164-8-124
- McWilliam H., Valentin F., Goujon M., Li W., Narayanasamy M., Martin J., Miyar T., Lopez R., Web services at the European Bioinformatics Institute-2009. Nucleic Acid Research, 37: doi:10.1093/nar/gkp302
- Medzhitov R. 2007. Recognition of microorganisms and activation of the immune response. Nature, 449: 819-826
- Morris G.M., Lim-Wilby M. 2008. Molecular Docking. V: Methods in Molecular Biology. Vol. 443. Molecular Modeling of Proteins. Kukol A. (ed.). New Jersey, Humana Press: 365-382
- NCBI Resource Coordinators. 2014. Database resources of the National Center for Biotechnology Information. Nucleic Acid Research, 42, D: 7-17
- Nerren J. R., He H., Genovese K., Kogut M. H. 2010. Expression of the avian-specific toll-like receptor 15 in chicken heterophils is mediated by Gram-negative and Gram-positive bacteria but not TLR agonists. Veterinary Immunology and Immunopathology, 136, 1-2: 151-156
- O'Neil L. A. J., Golenbock D., Bowie A. G. 2013. The history of Toll-like receptors – redefining innate immunity. Nature reviews Immunology, 13: 453-460
- Ramsden J. 2009. Bioinformatics. An Introduction. 2<sup>nd</sup> ed. London, Springer: 272 str.
- Schäffer A. A., Yi W., Wolf Y. I., Ponting C. P., Koonin E. V., Aravind L., Altschul S. F. 1999. IMPALA: matching a protein sequence against a collection of PSI-BLAST-constructed position-specific score matrices. Bioinformatics, 12: 1000-1011
- Sievers F., Wilm A., Dineen D., Gibson T. J. Karplus K., Li W., Lopez R., McWilliam H., Remmert M., Söding J., Thompson J. D., Higgins D. G. 2011. Fast, scalable generation of high-quality protein multiple sequence alignment using Clustal Omega. Molecular System Biology, 7: doi:10.1038/msb.2011.75
- Söding J., Biegert A., Lupas A. N. 2005. The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Research, 33: 244-248
- Suzek B. E., Huang H., McGarvey P., Mazumder R., Wu C. H. 2007. UniRef: comprehensive and non-redundant UniProt reference clusters. Bioinformatics, 23, 10: 1282-1288
- UniProt Consortium. 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acid Research, 42, D: 191-198
- UniProt Consortium. 2008. The Universal Protein Resource (UniProt). Nucleic Acid Research, 36, D: 190-195

- Venclovas Č. 2012. Methods for Sequence-Structure Alignment. V: Methods in Molecular Biology. Vol. 857. Homology Modeling: Methods and Protocols. Orry J.W. A., Abagyan R. (eds.). New Jersey, Humana Press: 55-82
- Wass M. N., Kelly L. A., Sternberg M. J. E. 2010. 3DLigandSite: predicting ligand-binding sites using similar structures. *Nucleic Acid Research*, 38: 469-473
- Xie Z. R., Hwang M. J. 2012. Ligand Binding Site Prediction Using Ligand Interacting and Binding Site-Enriched Protein Triangles. *Bioinformatics*, 28, 12: 1579-1585
- Yang J., Roy A., Zhang Y. 2013. Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment. *Bioinformatics*, 29, 20: 2588-2595

## **ZAHVALA**

Zahvaljujem se svoji mentorici prof. dr. Mojci Narat in somentorici dr. Ireni Oven, ki sta mi s temo o modeliranju omogočili, da se ukvarjam in seznanmi s področjem, ki me zelo zanima ter poglobim znanje bioinformatike. Zahvaljujem se jima za vse predloge in pomoč pri pisanku magistrske naloge.

Zahvaljujem se vsem prijateljem, ki so me podpirali pri mojih idejah in na ta način motivirali. Posebej bi se zahvalila Timoteju Lazarju, ki mi je s svojim znanjem računalništva pomagal pri razumevanju informacijskega dela in mi pri tem olajšal pisanje naloge.

Zahvaljujem se najbližnjim za vso pomoč, spodbudo, podporo in zaupanje, ki so mi ga nudili v času študija in pri oblikovanju magistrske naloge.

## PRILOGE

Priloga A: Preglednica šablon uporabljenih za modeliranje strukture molekule TLR 15 (PDB ID šablone, strežnik, ki je uporabil šablono za modeliranje, ime molekule šablone, domena molekule, ki je bila izbrana kot šablona ter povezava do molekule v podatkovni zbirki PDB)

PDB ID	Strežnik	Molekula	Domena	Povezava
1D0B_A	ModWeb	Internalin B	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1d0b">http://www.rcsb.org/pdb/explore/explore.do?structureId=1d0b</a>
1JL5_A	SWISS-MODEL	YopM	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1jl5">http://www.rcsb.org/pdb/explore/explore.do?structureId=1jl5</a>
1FYX_A	HHpred, Phyre <sup>2</sup>	TLR2	TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1fyx">http://www.rcsb.org/pdb/explore/explore.do?structureId=1fyx</a>
1FYV_A	MosWeb, HHpred, Phyre <sup>2</sup> , SWISS-MODEL	TLR1	TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1fyv">http://www.rcsb.org/pdb/explore/explore.do?structureId=1fyv</a>
1T3G_B	Phyre <sup>2</sup>	IL-1RAPL	TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1t3g">http://www.rcsb.org/pdb/explore/explore.do?structureId=1t3g</a>
1ZIW_A	HHpred, Phyre <sup>2</sup>	TLR3	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=1ziw">http://www.rcsb.org/pdb/explore/explore.do?structureId=1ziw</a>
2A0Z_A	Phyre <sup>2</sup>	TLR3	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2a0z">http://www.rcsb.org/pdb/explore/explore.do?structureId=2a0z</a>
2FT3_A	HHpred	Biglikan	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2ft3">http://www.rcsb.org/pdb/explore/explore.do?structureId=2ft3</a>
2ID5_D	Phyre <sup>2</sup>	Lingo-1	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2id5">http://www.rcsb.org/pdb/explore/explore.do?structureId=2id5</a>
2J67_A	HHpred	TLR10	TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2j67">http://www.rcsb.org/pdb/explore/explore.do?structureId=2j67</a>
2O6Q_A	SWISS-MODEL	VLR A	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2o6q">http://www.rcsb.org/pdb/explore/explore.do?structureId=2o6q</a>
2R9U_A	ModWeb	VLR 2913	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2r9u">http://www.rcsb.org/pdb/explore/explore.do?structureId=2r9u</a>
2V9T_B	SWISS-MODEL	Slit2	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2v9t">http://www.rcsb.org/pdb/explore/explore.do?structureId=2v9t</a>
2Z63_A	Phyre <sup>2</sup>	TLR4	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2z63">http://www.rcsb.org/pdb/explore/explore.do?structureId=2z63</a>
2Z64_A	Phyre <sup>2</sup>	TLR4	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2z64">http://www.rcsb.org/pdb/explore/explore.do?structureId=2z64</a>
2Z66_A	HHpred	TLR4	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2z66">http://www.rcsb.org/pdb/explore/explore.do?structureId=2z66</a>
2Z7X_B	Phyre <sup>2</sup>	TLR1	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2z7x">http://www.rcsb.org/pdb/explore/explore.do?structureId=2z7x</a>
2Z81_A	Phyre <sup>2</sup>	TLR2	EC-TLR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=2z81">http://www.rcsb.org/pdb/explore/explore.do?structureId=2z81</a>
3A79_A	Phyre <sup>2</sup>	TLR2	EC-TLR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3a79">http://www.rcsb.org/pdb/explore/explore.do?structureId=3a79</a>
3A79_B	Phyre <sup>2</sup>	TLR6	EC-TLR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3a79">http://www.rcsb.org/pdb/explore/explore.do?structureId=3a79</a>
3FXI_A	Phyre <sup>2</sup>	TLR4	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3fxi">http://www.rcsb.org/pdb/explore/explore.do?structureId=3fxi</a>
3G06_A	SWISS-MODEL	SspH2	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3g06">http://www.rcsb.org/pdb/explore/explore.do?structureId=3g06</a>
3J0A_A	HHpred, SWISS-MODEL, Phyre <sup>2</sup>	TLR5	LRR+TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3j0a">http://www.rcsb.org/pdb/explore/explore.do?structureId=3j0a</a>
3J0A_B	Phyre <sup>2</sup>	TLR5	LRR+TIR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3j0a">http://www.rcsb.org/pdb/explore/explore.do?structureId=3j0a</a>
3RG1_B	Phyre <sup>2</sup>	CD180	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3rg1">http://www.rcsb.org/pdb/explore/explore.do?structureId=3rg1</a>
3V47_B	Phyre <sup>2</sup>	TLR5	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3v47">http://www.rcsb.org/pdb/explore/explore.do?structureId=3v47</a>

3W3G_A	SWISS-MODEL	TLR8	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3W3G">http://www.rcsb.org/pdb/explore/explore.do?structureId=3W3G</a>
3W3J_B	Phyre <sup>2</sup>	TLR8	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3w3j">http://www.rcsb.org/pdb/explore/explore.do?structureId=3w3j</a>
3ZYI_A	HHpred, SWISS-MODEL	LRRC4	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=3ZYI">http://www.rcsb.org/pdb/explore/explore.do?structureId=3ZYI</a>
4G8A_B	HHpred, SWISS-MODEL, Phyre <sup>2</sup>	TLR4	EC-LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=4G8A">http://www.rcsb.org/pdb/explore/explore.do?structureId=4G8A</a>
4LSX_B	Phyre <sup>2</sup>	AtBAK1	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=4lsx">http://www.rcsb.org/pdb/explore/explore.do?structureId=4lsx</a>
4MN8_A	Phyre <sup>2</sup>	FLS2	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=4mn8">http://www.rcsb.org/pdb/explore/explore.do?structureId=4mn8</a>
4KT1_A	ModWeb	LGR4	LRR	<a href="http://www.rcsb.org/pdb/explore/explore.do?structureId=4kt1">http://www.rcsb.org/pdb/explore/explore.do?structureId=4kt1</a>